Two-Phase Behaviour in a Sequence of Random Variables

Pierre Abraham Mulamba Mutombo

Thesis presented in partial fulfilment
of the requirements for the degree of
Master of Science
at the University of Stellenbosch

Prof. A. E. Krzesinski
March 2007

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and has not previously in its entirety or in part been submitted at any university for a degree.

Signature .... Date ....

Abstract

Buying and selling in financial markets are driven by demand. The demand can be quantified by the imbalance in the number of shares Q_B and Q_S transacted by buyers and sellers respectively over a given time interval Δt. The demand in an interval Δt is given by $\Omega(t) = Q_B - Q_S$. The local noise intensity is given by $\Psi = \langle |a_i q_i - \langle a_i q_i \rangle| \rangle$ where $i = 1, \ldots, N$ labels the transactions in Δt, q_i is the number of shares traded in transaction i, $a_i = \pm 1$ denotes buyer- initiated and seller- initiated trades respectively and $\langle \cdots \rangle$ is the local expectation value computed from all the transactions during the interval Δt.

In a paper [1] based on data from the New York Stock Exchange Trade and Quote database during the period 1995-1996, Plerou, Gopikrishnan and Stanley [1] reported that the analysis of the probability distribution $P(\Omega|\Psi)$ of demand conditioned on the local noise intensity Ψ revealed the surprising existence of a critical threshold Ψ_c. For $\Psi < \Psi_c$, the most probable value of demand is roughly zero; they interpreted this as an equilibrium phase in which neither buying nor selling predominates. For $\Psi > \Psi_c$ two most probable values emerge that are symmetrical around zero demand, corresponding to excess demand and excess supply; they interpreted this as an out-of-equilibrium phase in which the market behaviour is buying for half of the time, and selling for the other half.

It was suggested [1] that the two-phase behaviour indicates a link between the dynamics of a financial market with many interacting participants and the phenomenon of phase transitions that occurs in physical systems with many interacting units.

This thesis reproduces the two-phase behaviour by means of experiments using sequences of random variables. We reproduce the two-phase behaviour based on correlated and uncorrelatd data. We use a Markov modulated Bernoulli process to model the transactions

and investigate a simple interpretation of the two-phase behaviour. We sample data from heavy-tailed distributions and reproduce the two-phase behaviour.

Our experiments show that the results presented in [1] do not provide evidence for the presence of complex phenomena in a trading market; the results are a consequence of the sampling method employed.

Opsomming

Aankope en verkope in finansiële markte word deur aanvraag gedryf. Aanvraag kan gekwantifiseer word in terme van die ongebalanseerdheid in die getal aandele Q_B en Q_B soos onderskeidelik verhandel deur kopers en verkopers in 'n gegewe tyd-interval Δt. Die aanvraag in 'n interval Δt word gegee deur $\Omega(t) = Q_B - Q_S$. Die lokale geraasintensiteit word gegee deur $\Psi = \langle |a_i q_i - \langle a_i q_i \rangle| \rangle$ waar $i = 1, \ldots, N$ die transaksies in Δt benoem, q_i die getal aandele verhandel in transaksies verwys, en $\langle \cdots \rangle$ op die lokale verwagte waarde dui, bereken van al die tansaksies tydens die interval Δt.

In 'n referaat [1] wat op data van die New York Effektebeurs se Trade and Quote databasis in die periode tussen 1995 en 1996 geskoei was, het Plerou, Gopikrishnan en Stanley [1] gerapporteer dat 'n analise van die waarskynlikheidsverspreiding $P(\Omega|\Psi)$ van aanvraag gekondisioneer op die lokale geraasintensiteit Ψ, die verrassende bestaan van 'n kritieke drempelwaarde Ψ_c na vore bring. Vir $\Psi < \Psi_c$ is die mees waarskynlike aanvraagwaarde nagenoeg nul; hulle het dit geïnterpreteer as 'n ekwilibriumfase waartydens nòg aankope nòg verkope die oormag het. Vir $\Psi > \Psi_c$ is die twee mees waarskynlike aanvraagwaardes wat te voorskyn kom simmetries rondom nul aanvraag, wat oorenstem met 'n oormaat aanvraag en 'n oormaat aanbod; hulle het dit geinterpreteer as 'n buite-ewewigfase waartydens die markgedrag die helfte van die tyd koop en die anderhelfte verkoop.

Daar is voorgestel [1] dat die tweefase gedrag op 'n verband tussen die dinamiek van 'n finansiele mark met baie deelnemende partye, en die verskynsel van fase-oorgange wat in fisieke sisteme met baie wisselwerkende eenhede voorkom, dui.

Hierdie tesis reproduseer die tweefase gedrag deur middel van eksperimente wat gebruik maak van reekse van lukrake veranderlikes. Ons reproduseer die tweefase gedrag gebaseer op gekorreleerde en ongekorreleerde data. Ons gebruik 'n Markov-gemoduleerde Bernoulli proses om die transaksies te moduleer en ondersoek 'n eenvoudige interpretasie van die tweefase gedrag.

Ons seem steekproefdata van "heavy-tailed" verspreidings en reproduseer die tweefase gedrag.

Ons ekperimente wys dat die resultate in [1] voorgested is nie bewys lewer vir die teenwoordigheid van komplekse verskynsel in'n handelsmark nie; die resultate is as gevolg van die metode wat gebruik is vir die generering van die steekproefdata.

Dedication

To my father S.Mulamba Munusadidi and my mother Charlotte Ngalula.

Acknowledgements

I would like to express my gratitude to my supervisor Prof A. E. Krzesinski for his encouragement to me throughout the period of this work; without his support, this work would not exist in its present form. I am also greatly indebted to the Department of Science and Technology(DST), Telkom SA, the African Institute for Mathematical Sciences (AIMS) and the Faculty of Science, University of Stellenbosch for providing me with the funds for my studies. I would like to thank the staff of the Department of Computer Science, University of Stellenbosch, especially Prof A. E. Krzesinski and Mr A. Bagula for creating an enabling environment for me to study in the department.

Many thanks to my wonderful parents – S. Mulamba and C. Ngalula – for allowing themselves to be used by God to influence my life a great deal. Finally, my family, friends, and colleagues deserve a special place in my heart for their support at all times.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the last few years there has been a surge of activity within the physics community in the emerging field of Econophysics – the study of economic systems from a physicist's perspective. Physicists tend to take a different view from economists and other social scientists, being interested in topics such as phase transitions and fluctuations.

In this thesis we use market terminology as an analogy in order to use the concepts of demand and local noise as used in the financial market literature.

The work presented in this thesis shows that the two-phase behaviour which was encountered in financial markets by analyzing the probability distribution of demand conditioned on its local noise intensity using the NYSE data can be obtained by sampling data from certain probability distributions. Here, we reproduce the two-phase behaviour by means of experiments using sequences of random variables.

1.1 Financial market, phase transition, demand

In economics, a financial market [2] is a mechanism which allows people to trade money for securities or commodities. Financial markets are affected by forces of supply and demand.

In physics, a phase transition [3] or phase change is the transformation of a thermodynamic system from one phase to another.

Demand [7] is the relationship, expressing different amounts of a product buyers are willing and able to buy at possible prices, assuming all other non-price factors remain the same.

Let Q_B and Q_S denote respectively the number of shares traded in buyer-initiated and seller-initiated transactions. The demand in an interval Δt is quantified by

$$\Omega(t) \quad = \quad Q_B - Q_S = \sum_{i=1}^{n} a_i q_i$$

where $i = 1, \ldots, N$ labels the transactions in an interval Δt, q_i is the number of shares traded in transaction i and $a_i = \pm 1$ denotes buyer- and seller- initiated trades respectively.

The local noise intensity in an interval Δt is given by the mean absolute deviation

$$\Psi(t) \quad = \quad \langle |a_i q_i - \langle a_i q_i \rangle| \rangle.$$

1.2 Problem statement

The first issue addressed in this thesis is to show that the two-phase behaviour can be reproduced using sequences of random variables.

The second isssue is to reproduce the two-phase behaviour based on correlated and uncorrelated data and investigate the effect of the correlation parameter. We use a Markov modulated Bernoulli process to model the transactions and investigate a simple interpretation of the two-phase behaviour where one share is bought or sold in each transaction.

The last issue addressed in this thesis is to reproduce the two-phase behaviour using data sampled from heavy tailed probability distributions, such as the Weibull and the Pareto distributions.

1.3 Thesis layout

This thesis is organized as follows:

Chapter 2 surveys the phenomena of phase transitions and gives a short literature survey on phase behaviour of financial markets. First, we introduce phase transitions in physical systems. Second, we present different classifications of phase transitions (such as first- and second- order phase transitions). Third, we present properties of phase transitions – critical points, symmetry, critical exponents and universality classes. Finally, we give a literature survey on phase behaviour of financial markets. Chapter 2 introduces and motivates the issues to be explored in chapters 4 and 5.

Chapter 3 presents some mathematical background used in this thesis. Before tackling the issues of two-phase behaviour reproduced in subsequent chapters, this chapter begins by presenting the concept of random variables. The concept of probability distribution is presented next. Finally, the inverse transform technique is presented.

Chapter 4 reproduces the two-phase behaviour using data sampled from a sequence of random variables. We begin by using a sequence of random variables sampled from a Markov modulated Bernoulli process, including samples from correlated and uncorrelated Markov modulated Bernoulli processes. We then perform experiments with a variant of a Markov modulated Bernoulli process. Finally, we consider the case of intervals of variable length that we take into account when reproducing the two-phase behaviour.

Chapter 5 reproduces the two-phase behaviour using data sampled from the Weibull and Pareto distributions. This includes experiments with different interval lengths and shape parameters of the probability distributions considered.

Chapter 6 presents a summary of the thesis.

Chapter 2

A survey of phase transitions

2.1 Phase transitions in physical systems

First, we present a brief description of the phenomena of phase transitions in physical systems and then we present a literature survey of phase behaviour of financial markets.

Definition

Consider a system with states X in contact with a heat bath at temperature $T = 1/\beta$.

Consider the conditional probability distribution $P(X|\beta) = \frac{1}{Z(\beta)} exp(-\beta E(X))$ of X conditioned on the temperature β.

The partition function is

$$Z(\beta) = \sum_x exp(-\beta E(x)).$$

The function $Z(\beta)$ is a continuous function of β. The derivatives of $Z(\beta)$ with respect to β are continuous.

The inverse temperature β is interpreted as defining an exchange rate between entropy and energy. $1/\beta$ is the amount of energy that must be given to a heat bath to increase its entropy by one nat. The system will be affected by other parameters such as the volume of the box it is in, V, in which case Z is also a function of V, $Z(\beta, V)$.

Phase transitions correspond to values of β and V at which the derivatives of Z have discontinuities or divergences [4]. Only systems with an infinite number of states can show phase transitions.

Consider a parameter N describing the size of the system. Phase transitions may appear in the limit $N \to \infty$. In real systems $N \sim 10^{23}$ [4].

The values of β and V at which the derivatives of Z have discontinuities or divergences are called critical points. At critical points systems change their behaviour. The critical points mark phase transition from one state of matter to another.

Examples of Phase transitions

1. The melting of a three-dimensional solid [5].
2. The transitions between the solid, liquid, and gaseous phases, due to the effects of temperature and pressure [3].

Classification of phase transitions

Phase transitions are categorized into "first-order" and "continuous" transitions.

First-order phase transitions

In a first-order phase transition, there is a discontinuous change of one or more order-parameters [4]. An order-parameter is a scalar function of the state of the system.

In first-order phase transitions the distinct states on either side of the critical point coexist exactly at the critical point. However, the states have different properties from each other. Slightly away from the critical point, there is a unique phase whose properties are continuously connected to one of the co-existent phases at the critical point. In that case there is discontinuous behaviour in various thermodynamic quantities as we pass through the critical point from one stable phase to another [5].

First-order transitions are associated with a latent heat and "mixed-phase regimes". In a mixed-phase regime some parts of the system complete the transition and others not. Mixed-phase regimes are difficult to study due to their dynamics. Many phase transition are mixed-phase regimes. Examples of first-order phase transitions are the solid/liquid/gas transitions and Bose-Einstein condensation [6].

Continuous transitions

In continuous phase or second-order phase transitions all order-parameters change continuously [4]. Parameters known as critical exponents characterize the continuous phase transitions. This class of phase transitions have no latent heat and the absence of latent heat makes the continuous phase easier to study than first-order phase transition. Examples of continuous phase transitions are the ferromagnetic transition and the super-fluid transition [6].

2.2 Phase behaviour of financial market

Buying and selling in financial markets are driven by demand. The demand can be quantified by the imbalance in the number of shares Q_B and Q_S transacted by buyers and sellers respectively over a given time interval Δt. The demand in an interval Δt is given by $\Omega(t) = Q_B - Q_S$. The local noise intensity is given by $\Psi = \langle |a_i q_i - \langle a_i q_i \rangle| \rangle$ where $i = 1, \ldots, N$ labels the transactions in Δt, q_i is the number of shares traded in transaction i, $a_i = \pm 1$ denotes buyer-initiated and seller-initiated trades respectively and $\langle \cdots \rangle$ is the local expectation value computed from all the transactions during the interval Δt.

In a paper [1] based on data from the New York Stock Exchange Trade and Quote database during the period 1995-1996, Plerou, Gopikrishnan and Stanley [1] reported that the analysis of the probability distribution of demand $P(\Omega|\Psi)$ conditioned on its local noise intensity Ψ revealed the surprising existence of a critical threshold Ψ_c. For $\Psi < \Psi_c$, the most probable value of demand is roughly zero; they interpreted this as an equilibrium phase in which neither buying nor selling predominates. For $\Psi > \Psi_c$ two most probable values emerge that are symmetrical around zero demand, corresponding to excess demand and excess supply. They interpreted this as an out-of-equilibrium phase in which the market behaviour is buying for half of the time, and selling for the other half. It was suggested [1] that the two phase behaviour indicates a link between the dynamics of a financial market with many interacting participants and the phenomenon of phase transitions that occurs in physical systems with many interacting units.

Using the data of the New York stock market (NYSE) between the years 2001-2002, Kaushik Matia and Kazuko Yamasaki [8] examined the out-of-equilibrium phase reported by Plerou et al. [1] and found that the observed two phase phenomenon is an artifact of the definition of the control parameters coupled with the nature of the probability distribution function of the share volume. They reproduced the two phase behaviour by a simple simulation demonstrating the absence of any collective phenomenon. They reported some interesting statistical regularities of the demand fluctuation of the financial market.

M.Potters and J.-P.Bouchard [9] showed that this apparent phase transition reported by Plerou et. al. [1] is a consequence of the conditioning and exists even in the absence of any non trivial collective phenomenon.

S.Sinha and S.Raghavendra [10] presented a model of two-phase behaviour and argued that it arose from interactions in a local neighbourhood and adaptation and learning based on information about the effectiveness of past choices.

B.Zheng, T.Qiu and F.Ren [11] examined the German financial index DAX, minority games, and dynamic herding models. They observed that the two-phase phenomenon is an important characteristic of financial dynamics, independent of volatility clustering. An interacting herding model correctly produces the two-phase phenomenon.

M.Wyart and J.-P.Bouchaud [12] studied a generic model for self-referential behaviour in financial markets, where agents attempt to use some (possibly fictitious) causal correlations between a certain quantitative information and the price itself. This correlation is estimated using the past history, and is used by a fraction of the agents to devise active trading strategies. The impact of these strategies on the price modifies the observed correlations. A potentially unstable feedback loop appears and destabilizes the market from efficient behaviour. For large enough feedbacks, they found a “phase transition” beyond which non trivial correlations spontaneously set in, and where the market switches between two long lived states, which they called conventions. This mechanism leads to overreaction and excess volatility, which may be considerable in the convention phase. A particularly relevant case is when the source of information is the price itself. The two conventions correspond then to either a trend following regime or to a contrarian (mean reverting) regime. They provided some empirical evidence for the existence of these conventions in real markets.

F.F.Gong, F.X.Gong and F.Y.Gong [13] investigated the dynamic behaviour of financial markets with internal interactions between agents and with external "fields" from other systems using the approach of Grossman and Stiglitz [14] for inefficient markets, and Keynes for interference of the market using the physics of finance. The simulation results indicated that the NYSE data analyzed in [1] can be fitted by an equation of order parameter Φ and local deviation R of type:$-(R + 0.03)\Phi + 0.6\Phi^3 + 0.02 = 0$, which is shown to be in remarkable agreement with Plerou's data.

M.Forster and B.Halfpap [15] found that the existence of two distinct phase behaviour is a direct consequence of long-tailed distributions of independent random variables.

A. Costa, A.E. Krzesinski, M. Ramakrishnan and P.G. Taylor [16] urge caution with findings of [1]. In particular, they show that the statistical technique employed in [1] to analyse stock trading data also produces evidence of two-phase behaviour when used to analyse a sequence of independent and identically distributed random variables.

2.3 Chapter summary

This chapter gives a short survey of phase transitions in physics and presents different classification of phase transitions and a short literature survey of phase transitions of financial markets.

Chapter 3

Mathematical background

In this chapter we present some of the mathematical concepts used in the subsequent chapters. The basic concepts of random variables, the Bernoulli process and the Markov modulated Bernoulli process that we use in chapter 4 are reviewed. We next present some mathematical properties of the probability distributions that we use in chapter 5 and the inverse transform technique [17] that is used to generate samples from a given probability distribution.

3.1 Random variables

Definition

In practice, if the outcome of a process is not known in advance, then the process is nondeterministic or stochastic. A stochastic process is the set of outcomes of a random occurrence of a process, indexed by time. A state is the condition of a stochastic process, at a specific time, described by means of random variables. The state space $\mathcal{S}$ is the set of all possible states of a stochastic process. Most stochastic processes are expressed in terms of random variables: the number of jobs in the queue, the fraction of time a processor is busy, the amount of time a server is operational are examples of random variables. Random variables are functions that map a real value to every random outcome of the state space. Random variables that take a countable number of values are discrete, otherwise they are continuous.

A random variable is defined as a measurable function from a probability space to some measurable space.

Let (Ω, A, P) be a probability space. A function $X : \Omega \longrightarrow R$ is a random variable if for every subset $A_r = \{\omega : X(\omega) \leq r\}$ where $r \in R$, we also have $A_r \in A$.

The first item, Ω, is a nonempty set, whose elements are known as outcomes. An element of Ω is given the symbol ω.

The second item, A, is a set, whose elements are called events. The events are subsets of Ω. The set A has to be a $\sigma - algebra$. (Ω, A) forms a measurable space. An event is a set of outcomes.

Let X be a set. A $\sigma - algebra$ or $\sigma - field$ [19] $\mathcal{F}$ is a nonempty collection of subsets of X such that the following hold:

1. $X \in \mathcal{F}$.
2. if $E \in \mathcal{F}$ then $E^c \in \mathcal{F}$, where $E^c = X \backslash E$.
3. if $E_n \in \mathcal{F}$ for all $n = 1, 2, \cdots$ then $\bigcup E_n \in \mathcal{F}$.

A measurable space is a set considered together with the $\sigma - algebra$ on the set [20].

The third item, P, is called the probability measure, or the probability. It is a function from A to the real numbers, assigning each event a probability between 0 and 1. P must be a measure and $P(\Omega) = 1$.

Function of random variables

If we have a random variable X on Ω and a measurable function $f : R \longrightarrow R$, then $Y = f(X)$ will also be a random variable on Ω, since the composition of measurable function is also measurable. The cumulative distribution function of Y is $F_Y(y) = P(f(X) \leq y)$.

3.2 Bernoulli process

Definition

A Bernoulli process [18] is a discrete-time stochastic process consisting of a finite or infinite sequence of independent random variables X_1,X_2,X_3,$\cdots$ such that for each i, the value of X_i is either 0 or 1.

The Bernoulli process can be formalized in the language of probability spaces. A Bernoulli process is a probability space (Ω, A, P_r) together with a random variable X over the set $\{0, 1\}$, so that for every $\omega \in \Omega$, $X_i(\omega) = 1$ with probability p and $X_i(\omega) = 0$ with probability $1 - p$.

Bernoulli sequence

Given a Bernoulli process defined on a probability space (Ω, A, P_r), then associated with every $\omega \in \Omega$ is a sequence of integers $Z^\omega = \{n \in Z : X_n(\omega) = 1\}$ which is called a Bernoulli sequence [21].

3.3 Markov modulated Bernoulli process

The Markov modulated Bernoulli (MMB) [22] process is obtained by assuming that the success probability of a Bernoulli process evolves over time according to a Markov chain.

A discrete Markov chain [23] is defined as a collection of random variables (X_t) (where the index t runs through $0, 1, \cdots$) having the property that, given the present, the future is conditionally independent of the past. Thus

$$P(X_t = j|X_0 = x_0, X_1 = x_1, \ldots, X_{t-1} = x_{t-1}) = P(X_t = j|X_{t-1} = x_{t-1}).$$

If a Markov sequence of random variables X_n takes the discrete values $a_1, \ldots, a_N$, then

$$P(X_n = a_{i_n}|X_{n-1} = a_{i_{n-1}}, \ldots, X_1 = a_{i_1}) = P(X_n = a_{i_n}|X_{n-1} = a_{i_{n-1}})$$

and the sequence $\{X_n\}$ is a Markov chain.

3.4 Probability distribution

In this section we present two probability distributions which are used in chapter 5 of this thesis.

3.4.1 Discrete distributions

A discrete probability function $p(x)$ satisfies the following properties

1. The probability that x can take a specific value is $p(x)$. That is $P[X = x] = p(x) = p_x$.
2. $p(x)$ is non-negative for all real x.
3. $\sum_j p_j = 1$ where j represents all possible values that x can have and $p_j = p(x_j)$.

One consequence of properties 2 and 3 is that $0 \leq p(x) \leq 1$.

3.4.2 Continuous distributions

A continuous probability function $f(x)$ satisfies the following properties

1. The probability that x is between two points a and b is $p[a \leq x \leq b] = \int_a^b f(x)dx$.
2. $f(x)$ is non-negative for all real x.
3. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Discrete probability functions are referred to as probability mass functions and continuous probability functions are referred to as probability density functions. The term probability functions covers both discrete and continuous distributions.

3.4.3 Heavy-tailed and power-tailed distributions

A distribution is said to be heavy-tailed [24] if

$$P[X > x] \sim x^{-\alpha}, \text{ as } x \rightarrow \infty.$$

The probability mass function of a heavy-tailed distribution is given by

$$p(x) = \alpha k^{\alpha} x^{-\alpha-1}, \ \alpha, k > 0, \ x \geq k$$

and its cumulative distribution function is given by

$$F(x) = P[X \leq x] = 1 - (\frac{k}{x})^{\alpha}$$

where k represents the smallest value the random variable can take.

A distribution is power-tailed [24] if its tail decays as a power law. A power law relationship between two scalar quantities x and y is one where the relationship can be written as

$$y = ax^k$$

where a is the constant of proportionality and k the exponent of the power law.

The Pareto distribution is a classic case of a distribution exhibiting power-tailed behavior in the entire range of its parameters. The Weibull distribution is heavy-tailed, but not power-tailed.

We consider the heavy-tailed Weibull and Pareto distribution in our experiments.

The Weibull distribution

The Weibull distribution is a special case of the Generalized Extreme Value distribution. It has been extensively used as a model of time to failure of manufactured items and has become one of the principal tools of reliability engineering. The applications of the Weibull distribution also include finance and climatology. The distribution is named after the Swedish engineer Wallodi Weibull.

The Weibull distribution is most commonly used in life data analysis, though it has found other applications as well. The Weibull distribution is often used in place of the normal distribution due to the fact that a Weibull variate can be generated through inversion, while normal variates are generated using more complicated methods. Weibull distributions may also be used to represent manufacturing and delivery times in industrial engineering problems, while it is very important in extreme value theory and weather forecasting. It is also a popular statistical model in reliability engineering and failure analysis, while it is widely applied in radar systems to model the dispersion of the received signal level produced by some types of clutters.

The Weibull probability distribution is characterized by location, scale and shape parameters. The location parameter shifts the distribution left or right on the horizontal axis. The scale parameter defines the range and a practical maximum, also known as the characteristic life. The shape parameter determines the profile of the distribution. The range of the Weibull distribution is $[0, +\infty)$.

Functions

The Weibull probability density function is

$$P(x) \;=\; \frac{\gamma}{\alpha}\Big(\frac{x-\mu}{\alpha}\Big)^{\gamma-1}\exp\left(-\Big(\frac{x-\mu}{\alpha}\Big)^{\gamma}\right)$$

for $x \geq \mu$ and $\alpha > 0$ where γ is the shape parameter, μ is the location parameter and α is the scale parameter. The case $\mu = 0$ and $\alpha = 1$ gives the standard Weibull probability density function

$$P(x) \;=\; \gamma x^{\gamma-1}\exp(-x^{\gamma})$$

for $\gamma > 0$.

The Weibull distribution function is

$$D(x) \;=\; 1 - e^{-(x/\alpha)^{\gamma}}$$

for $x \in [0, \infty)$.

Properties

The mean, variance, skewness, and kurtosis of the Weibull distribution are

$$\begin{aligned}
\mu &= \alpha\Gamma(1+\gamma^{-1}) \\
\sigma^2 &= \alpha^2(\Gamma(1+2\gamma^{-1}) - \Gamma^2(1+\gamma^{-1})) \\
\gamma_1 &= \frac{2\Gamma^3(1+\gamma^{-1}) - 3\Gamma(1+\gamma^{-1})\Gamma(1+2\gamma^{-1})}{(\Gamma(1+2\gamma^{-1}) - \Gamma^2(1+\gamma^{-1}))^{3/2}} + \frac{\Gamma(1+3\gamma^{-1}))}{(\Gamma(1+2\gamma^{-1}) - \Gamma^2(1+\gamma^{-1}))^{3/2}} \\
\gamma_2 &= \frac{f(\gamma)}{(\Gamma(1+2\gamma^{-1}) - \Gamma^2(1+\gamma^{-1}))^2}.
\end{aligned}$$

where $\Gamma(z)$ is the Gamma function and

$$f(\gamma) = -6\Gamma^4(1+\gamma^{-1}) + 12\Gamma^2(1+\gamma^{-1})\Gamma(1+2\gamma^{-1}) - 3\Gamma^2(1+2\gamma^{-1}) - 4\Gamma(1+\gamma^{-1})\Gamma(1+3\gamma^{-1}) + \Gamma(1+4\gamma^{-1})$$

The Gamma function is defined by $\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt$.

The Pareto distribution

The Pareto distribution is a highly left skewed distribution defined in terms of a mode and a shape factor. It is a heavy-tailed distribution meaning that a random variable sampled from a Pareto distribution can have extreme values.

Applications of the Pareto distribution include insurance, where it is used to model claims, where the minimum claim is also the modal value, but where there is no set maximum. In climatology it is used to describe the occurrence of extreme weather. The Pareto distribution has been proposed as a model for oil and gas discoveries where the minimum size is set by the economics of production.

The Pareto distribution was originally developed to describe the distribution of income, where a high proportion of the population have low income, whilst only a few people have very high incomes.

The mode value of the Pareto distribution is the minimum value. The shape parameter of the Pareto distribution determines the concentration of data towards the mode.

The range of random numbers generated from the Pareto distribution is from the mode to $+\infty$.

Functions

The Pareto probability density function is

$$P(x) = \gamma\beta^{\gamma}/x^{\gamma+1}$$

and its distribution function is

$$D(x) = 1 - (\beta/x)^{\gamma}$$

defined over the interval $x \geq \beta$ with $\gamma > 0$, where β and γ are the mode and the shape parameter respectively.

Properties

The mean, variance, skewness, and kurtosis of the Pareto distribution are

$$\begin{aligned}
\mu &= \frac{\gamma\beta}{\gamma-1} \\
\sigma^2 &= \frac{\gamma\beta^2}{(\gamma-1)^2(\gamma-2)} \\
\gamma_1 &= \frac{2(\gamma+1)}{\gamma-3}\sqrt{\frac{\gamma-2}{\gamma}} \\
\gamma_2 &= \frac{6(\gamma^3+\gamma^2-6\gamma-2)}{\gamma(\gamma-3)(\gamma-4)}.
\end{aligned}$$

3.5 Inverse Transform Technique

Inversion is a general method for sampling random variables. It makes use of the fact that the transformation $X = F^{-1}(U)$ gives a random variable X with distribution function F provided the inverse function F^{-1} exists. This is a simple consequence of the change of variables formula, this time with $g(U) = F^{-1}(U)$. Since $g^{-1}(x) = F(x)$, the density of X becomes $\frac{d}{dx}F(x) = f(x)$, which is the probability density corresponding to the distribution function F.

We use the inverse transform technique to generate a sample $\{X_i\}$ from given probability distribution.

Let $F(x) = Pr(X \leq x)$ denote the distribution function of the random variable X, and let $F^{-1}(\cdot)$ denote the inverse function of $F(\cdot)$. Thus if $F(x) = u$ then $x = F^{-1}(u)$.

Let $U \sim U[0,1]$, and let $X = F^{-1}(U)$. Then X has distribution function $F(\cdot)$. In fact,

$$Pr(X \leq x) = Pr(F^{-1}(U) \leq x) = Pr(U \leq F(x)) = F(x)$$

since $Pr(U \leq u) = u$. The inverse function $F^{-1}(\cdot)$ is well defined for continuous random variables. For discrete random variables $F^{-1}(\cdot)$ is defined by $F^{-1}(u) = \min\{x : F(x) \geq u\}$. To use the inversion method, the inverse function F^{-1} either has to be available explicitly, as in the exponential, Weibull, logistic and Pareto cases, or has to be computable in a reasonable amount of time. The equality $x = F^{-1}(u)$ is equivalent to $u = F(x)$, so that finding x for given u is equivalent to finding a root of the equation $F(x) - u = 0$. When F is strictly monotone, there is only one root and standard numerical root-finding algorithms can be used, provided that $F(x)$ itself is easy to evaluate. If it is required to sample repeatedly from the same distribution, it may be worthwhile devoting some time to the development of an accurate approximation to F^{-1} beforehand.

The main advantage of the inversion method is that generally it is easy to verify that a computer algorithm which using it, is written correctly. In this sense the method is efficient.

3.6 Chapter summary

This chapter briefly presents the mathematical concepts used in the following chapters. We start with a presentation of random variables. We then present the Bernoulli process and Markov modulated Bernoulli process. The fourth section presents probability distributions: discrete and continuous probability distributions, heavy- and power- tailed distributions, the Weibull and Pareto distributions. Finally, we present the inverse transform technique.

Chapter 4

Two-phase behaviour in a sequence of random variables

4.1 Introduction

In this chapter we reproduce two-phase behaviour among sequences of correlated and uncorrelated random variables.

We use market terminology as an analogy so that we can use the concepts of demand and local noise as used in financial market models. Our experiments are not based on real financial data but are based on random variables.

Our experiment to reproduce the two-phase behaviour is as follows

1. We reproduce two-phase behaviour in a sequence of correlated Bernoulli random variables and investigate how it depends on the value of the correlation parameter.
2. We reproduce two-phase behaviour among uncorrelated sequences of normally distributed random variables and investigate how it depends on the parameter of the normal distribution.
3. We reproduce two-phase behaviour among correlated sequences of normally distributed random variables and investigate how it depends on the correlation parameter.

In the next section we consider a Bernoulli process and a Markov modulated Bernoulli process and reproduce two-phase behaviour. We present a simple interpretation of the origin of the two phase behaviour among sequences of such random variables.

We will generate a sequence of independent random variables $X_1,X_2,X_3,\cdots,X_N$ such that for each i, the value of X_i is either -1 or $+1$.

Let Q_B and Q_S respectively denote the number of shares traded in buyer-initiated and seller-initiated transactions. The demand in an interval Δt is quantified by

$$\Omega(t) \;=\; Q_B - Q_S = \sum_{i=1}^{n} a_i q_i$$

where $i = 1,\cdots,N$ labels the transactions in an interval Δt, q_i is the number of shares traded in transaction i and $a_i = \pm 1$ denotes buyer- and seller- initiated trades respectively.

The local noise intensity in an interval Δt is given by the mean absolute deviation

$$\Psi(t) \;=\; \langle |a_i q_i - \langle a_i q_i \rangle| \rangle.$$

Consider a sequence of N random variables $(a_i)_{i=1,\cdots,N}$ where $a_i = \pm 1$. Let n of the N random variables each have value 1. Therefore the remaining $N - n$ of the N random variables each have value -1. The demand Ω is

$$\Omega \;=\; \sum_{i=1}^{n} a_i = \underbrace{1 + 1 \cdots + 1}_{n} + \underbrace{(-1 - 1 \cdots - 1)}_{N-n} = n - (N-n) = 2n - N.$$

The average value of the $(a_i)_{i=1,\cdots,N}$ is

$$\omega \;=\; \frac{\sum_{i=1}^{n} a_i}{N} = \frac{2n - N}{N} = \frac{\Omega}{N}.$$

Let

$$\begin{aligned}\Psi \;&=\; \frac{1}{N}\sum_{i=1}^{n} |a_i - \omega| = \frac{1}{N}(\underbrace{|1-\omega| + |1-\omega| \cdots + |1-\omega|}_{n} + \underbrace{|-1-\omega| + |-1-\omega| \cdots |-1-\omega|}_{N-n}) \\ &=\; \frac{1}{N}(n|1-\omega| + (N-n)|-1-\omega|) = \frac{1}{N}(n(1-\omega) + (N-n)(1+\omega)) = 1 - \omega^2\end{aligned}$$

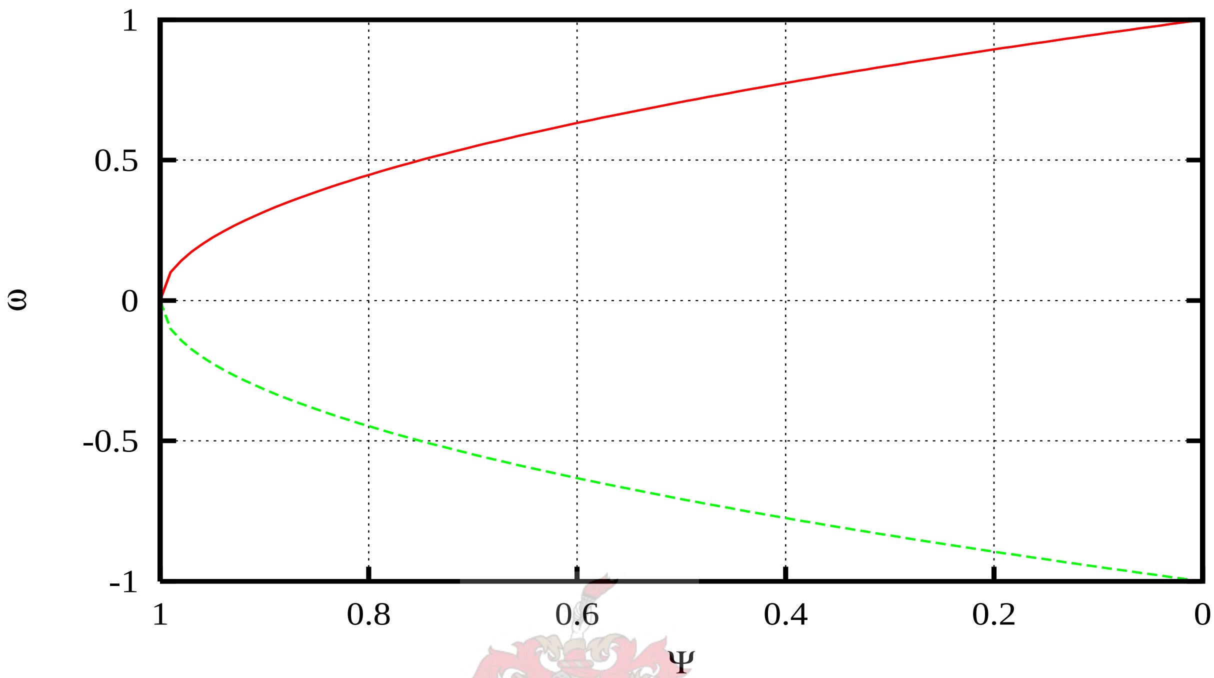


Figure 4.1: *ω as a function of Ψ*

denote the mean absolute deviation of the $(a_i)_{i=1,\cdots,N}$. The variance σ^2 of the $(a_i)_{i=1,\cdots,N}$ is

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{n}(a_i-\omega)^2 = \frac{N}{N-1}(1-\omega)^2.$$

The equation $\Psi = 1-\omega^2$ yields $\omega = \sqrt{1-\Psi}$. For $0 \leq \Psi < 1$, we have $\omega = \pm\sqrt{1-\Psi}$ and probability density $P(\Omega|\Psi)$ is bi-modal. For $\Psi = 1$, we have $\omega = 0$ and the probability density $P(\Omega|\Psi)$ is uni-modal. Fig.4.1 shows ω as a function of Ψ.

4.2 The Markov modulated Bernoulli process

A sequence $(a_i)_{i=1,\dots}$ of random variables is sampled from a Markov modulated Bernoulli (MMB) distribution and a correlation is introduced among the random variables (a_i). We use algorithm 1 to generate a sequence of MMB random variables. We vary the correlation parameter to reproduce two-phase behaviour among this correlated sequence of random variables.

Consider the MMB process

$$\begin{aligned} Pr(a_{i+1} = 1) &= APr(a_i = 1) + (1 - A)Pr(a_i = -1) \\ Pr(a_{i+1} = -1) &= 1 - Pr(a_{i+1} = 1) \end{aligned}$$

where A is the correlation parameter. Let $(a_i) = \pm 1$ where $i = 1, \cdots, N$ indicates the events in an interval of length N.

The interval length N is set to 10. The demand Ω and the local noise intensity Ψ are computed for each interval as follows:

$$\begin{aligned} \Omega &= 2n - N \\ \omega &= \frac{\Omega}{N} \\ \Psi &= 1 - \omega^2. \end{aligned}$$

Algorithm 1

```
a = 1
initialize( A ∈ (0, 1))
for i = 1 to M  do
   Z = U(0, 1)            //a standard uniform RV
   if Z > A then
      a = −a
   end if
end for
```

In the first experiment we set the correlation parameter A to 0.01. Fig.4.2(a) shows that the probability density $P(\Omega|\Psi)$ of the demand Ω conditioned on its local noise intensity

Ψ is essentially uni-modal, consisting of a single peak at $\Psi = 1$ and two small peaks at $\Psi = 0.96$.

Fig.4.2(b) represents a similar experiment except that the correlation parameter $A = 0.1$. In this case, two peaks emerge at $\Psi = 0.96$. The conditional probability density $P(\Omega|\Psi)$ is uni-modal at $\Psi = 1$ and bi-modal at $\Psi = 0.96$.

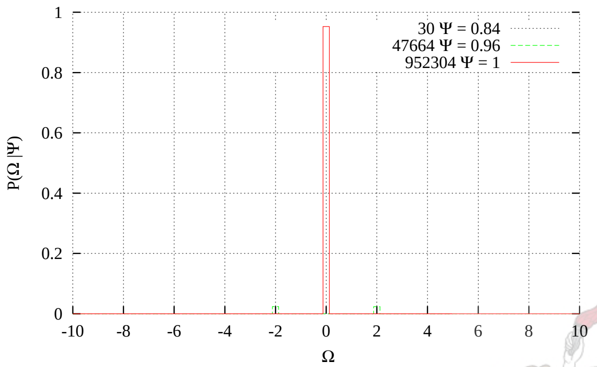
4.2.1 Two-phase behaviour among random variables sampled from uncorrelated MMB

Consider the correlation parameter $A = 0.5$ which produces a sequence of un-correlated Bernoulli random variables. Fig.4.2(c) shows that the probability density $P(\Omega|\Psi)$ is uni-modal at $\Psi = 1$ and bimodal at $\Psi = 0.96$, $\Psi = 0.84$, $\Psi = 0.64$ and $\Psi = 0.36$. Analyzing Fig.4.2(a), (b) and (c) we notice that more distinctive peaks emerge at different values of $\Psi < 1$ as we increase the correlation parameter.

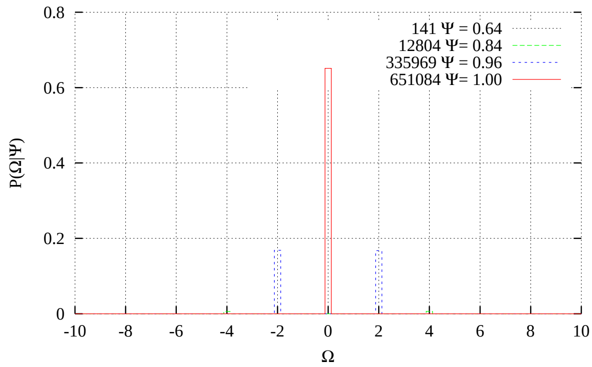
4.2.2 Large correlation parameter

Here we investigate a large value of the correlation parameter namely $A = 0.9$. Fig.4.2(d) shows that a large correlation parameter increases the probability density $P(\Omega|\Psi)$ at $\Psi < 1$ and decreases the probability density $P(\Omega|\Psi)$ at $\Psi = 1$. We have two large peaks at $\Omega = \pm 10$ and the other peaks have almost the same hight. The probability density $P(\Omega|\Psi)$ is bi-modal for values of $\Psi < 1$ and uni-modal at $\Psi = 1$.

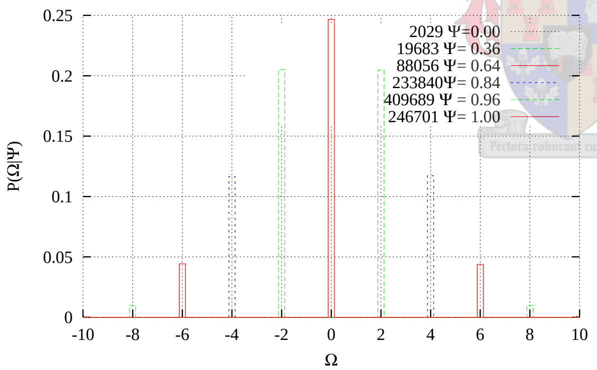
Algorithm 2 is used to compute the conditional probability density $P(\Omega|\Psi)$ of Ω given Ψ.



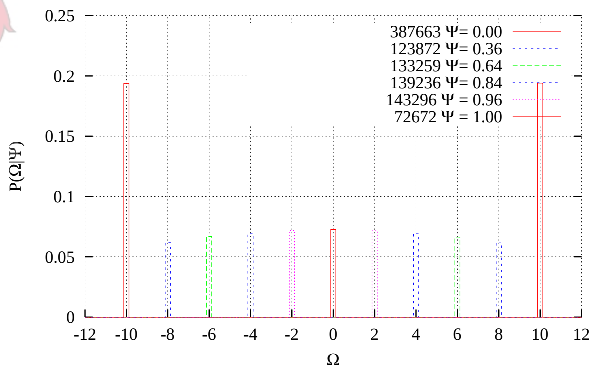
(a) *Correlation parameter* $A = 0.01$



(b) *Correlation parameter* $A = 0.1$



(c) *Correlation parameter* $A = 0.5$



(d) *Correlation parameter* $A = 0.9$

Figure 4.2: $P(\Omega|\Psi)$ *using the Markov modulated Bernoulli Process: interval length* $N = 10$.

Algorithm 2

Inputs: M transactions over K intervals each of length N transactions, a noise selection range $\Psi = [\Psi_-, \Psi_+)$, the histogram bin width b.

$m = 0$ // the number of bins $m = 0$
for $k = 1$ to K **do**
  compute Ω_k and Ψ_k
  if $\Psi_- \leq \Psi < \Psi_+$ **then**
    $i = \lfloor \Omega_k / b \rfloor$ // compute the bin index
    if $m < |i|$ **then**
      $m = i$ // update the number of bins
    end if
    $H(i|\Psi) = H(i|\Psi) + 1$ // update the histogram
    $K(\Psi) = K(\Psi) + 1$ // count the intervals in Ψ
  end if
end for
for $i = -m$ to m **do**
  $H(i|\Psi) = H(i|\Psi)/K(\Psi)$ // normalize the histogram H
  $H'(i|\Psi) = H(i|\Psi)/K$ // normalize the histogram H'
end for

4.3 A Simple market model

In this section we use a variant of the MMB process to model the behaviour of the trading of shares of a single stock and we find the existence of a critical threshold similar to that in [1].

Consider a simple market model for trading one stock. Let $a_i = \pm 1$ denote the buyer- and seller-initiated trades respectively where $i = 1, \cdots, N$ labels the trades in an interval of N transactions.

Consider a sequence $(a_i|q_i|)_{i=1,\ldots,N}$ where a_i is a random variable sampled from a MMB distribution and q_i is sampled from a normal distribution. Let $|q_i|$ model the number of shares bought or sold in the i-th transaction.

We use the same approach as in the previous section to reproduce two-phase behaviour. We start with a small value of the correlation parameter, then we consider un-correlated data and finally we set the value of the correlation parameter to a large value.

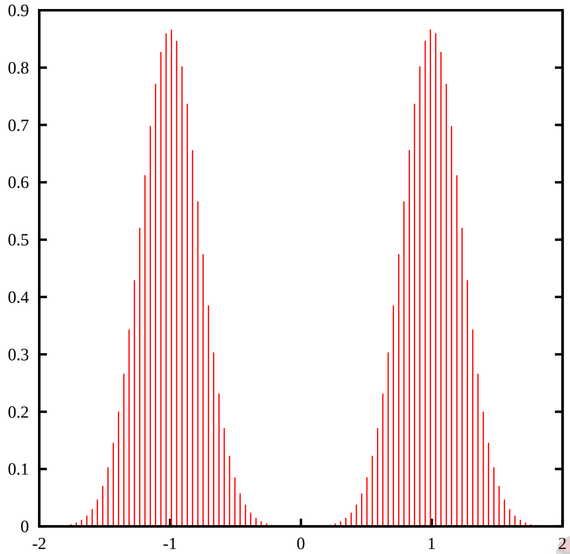
Algorithm 2 is used to compute the probability density $P(\Omega|\Psi)$ of the demand Ω conditioned on its local noise intensity Ψ. Algorithm 3 is used to generate a sequence of $(a_i|q_i|)_{i=1,\dots,M}$. Fig.4.3(a) through (d) show graphs of the sample $(a_i|q_i|)$ for different values of the mean and standard deviation of the normal distribution N.

Algorithm 3

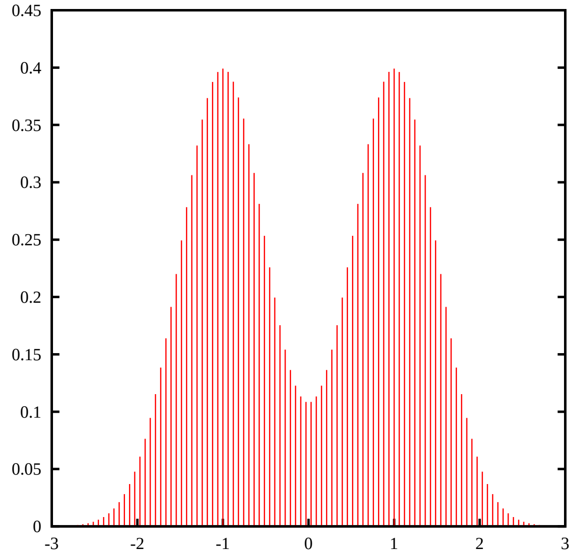
```
a = 1
initialize( A ∈ (0,1), σ ≥ 1, μ ≥ 0)
for i = 1 to M  do
  Z = U(0,1)              // a standard uniform RV
  if a = 1 then
    if Z > A then
      a = −1
    end if
  else
    if Z < (1 − A) then
      a = 1
    end if
  end if
  q = N(0,1)              // a standard normal RV
  q = σ ∗ N(0,1) + μ      // a normal RV
  q = a∗ | q |            // introduce the correlation
end for
```


4.3.1 Small value of the correlation parameter

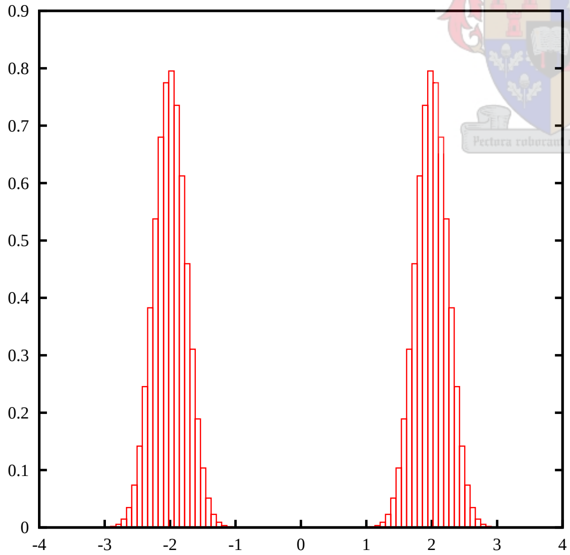
Fig.4.4 shows that the probability density$P(\Omega|\Psi)$ is uni-modal with a single peak centered at zero. The correlation parameter A is 0.05 and the interval length N is 10. We sample the random variables from two normal distributions $N(1, 0.25)$ and $N(-1, 0.25)$.



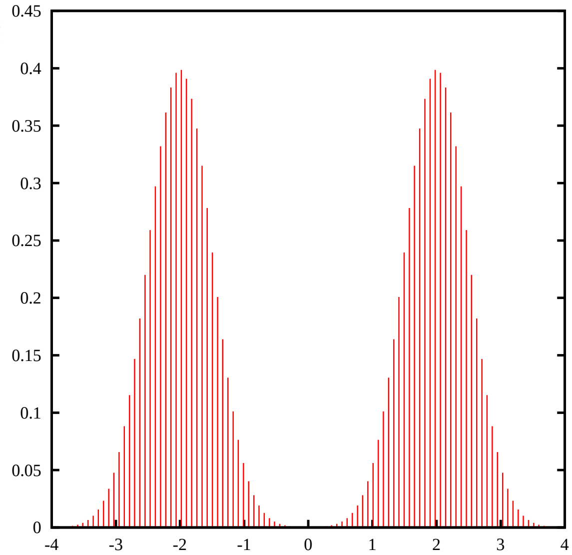
(a) (q_i) *sampled from* $N(\pm 1, 0.25)$ *and* (a_i) *sampled from a MMB*



(b) (q_i) *sampled from* $N(\pm 1, 0.5)$ *and* (a_i) *sampled from a MMB*



(c) (q_i) *sampled from* $N(\pm 2, 0.25)$ *and* (a_i) *sampled from a MMB*



(d) (q_i) *sampled from* $N(\pm 2, 0.5)$ *and* (a_i) *sampled from a MMB*

Figure 4.3: *Graphs of sample* $(a_i|q_i|)$.

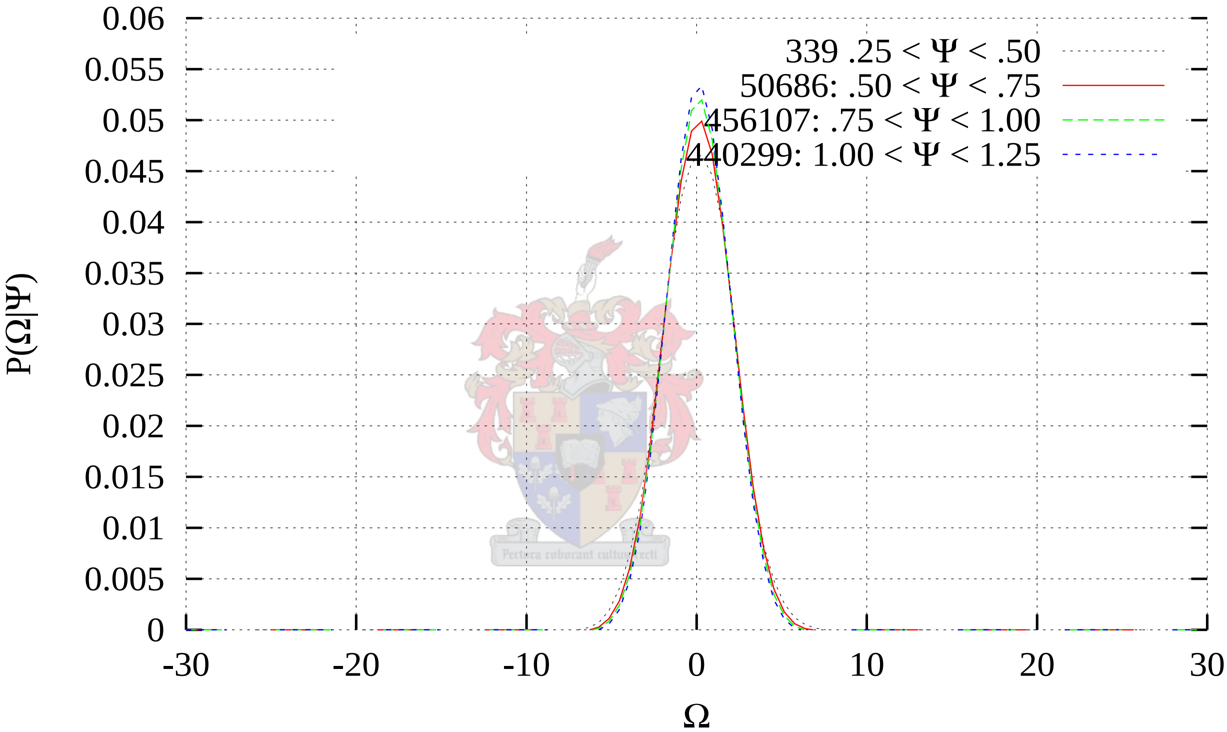


Figure 4.4: $P(\Omega|\Psi)$*: Correlation parameter* $A = 0.05$

4.3.2 Uncorrelated data

We perform experiments and reproduce the uni- and bi- modality of the probability density $P(\Omega|\Psi)$ by varying the mean and the variance of the normal distribution N while keeping the correlation parameter A fixed to 0.5. In other words, we perform experiments with un-correlated data.

The experiment entails three steps

1. Increase the correlation parameter A .

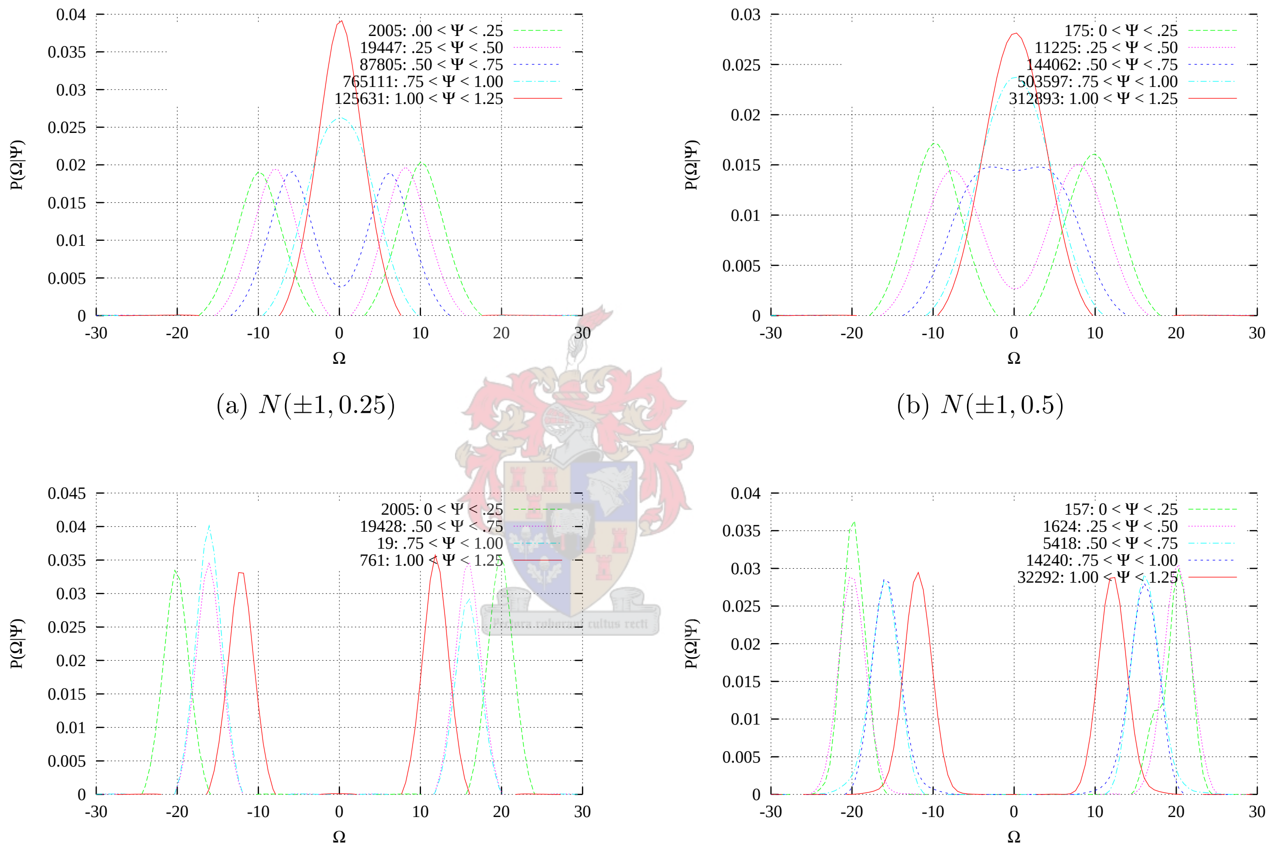
   We first set the correlation parameter A to 0.5 and perform a similar experiment as in the previous section except that the correlation parameter is 0.5. The approach is to see the impact of the correlation parameter in the experiment. We observe that the probability density $P(\Omega|\Psi)$ is uni- and bi- modal. We have a large peak centered at 0 for $1 < \Psi < 1.25$ and another peak centered at 0 for $0.75 < \Psi < 1$. For $0 < \Psi < 0.75$ the probability density is bi-modal as shown in Fig.4.5(a). Thus, increasing the correlation parameter causes the probability density $P(\Omega|\Psi)$ to become uni- and bi-modal. We note that the value 0.5 of the correlation parameter produces un-correlated data.

2. Change the standard deviation of the normal distribution N.

   We use un-correlated data (the correlation parameter A is 0.5) and we sample from $N(\pm 1, 0.5)$. Fig.4.4(b) shows graph of the sample used to perform the experiment and Fig.4.5(b) shows that the probability density $P(\Omega|\Psi)$ is uni-modal for $0.75 < \Psi < 1$ and $1 < \Psi < 1.25$ bi-modal for $0 < \Psi < 0.75$.

3. Change the mean and standard deviation of the normal distribution N.

   We sample from $N(\pm 2, 0.25)$ and $N(\pm 2, 0.5)$ as shown in Fig.4.4(c) and Fig.4.4(d) respectively. Fig.4.5(c) and (d) show the The probability density $P(\Omega|\Psi)$ is essentially bi-modal in both cases.



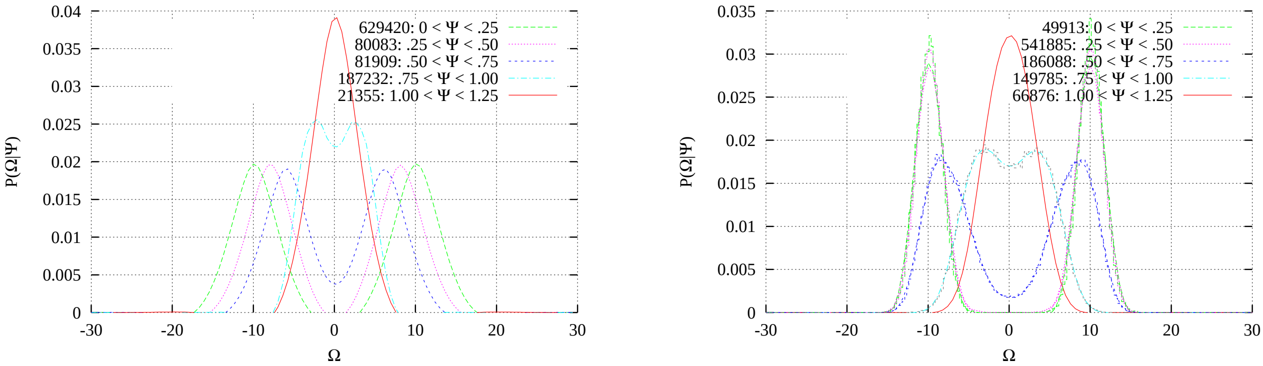
(a) $N(\pm 1, 0.25)$

(b) $N(\pm 1, 0.5)$

(c) $N(\pm 2, 0.25)$

(d) $N(\pm 2, 0.5)$

Figure 4.5: $P(\Omega|\Psi)$: *Correlation parameter* $A = 0.5$, *interval length* $N = 10$.



(a) $N(\pm 1, 0.25)$ (b) $N(\pm 1, 0.5)$

Figure 4.6: $P(\Omega|\Psi)$: *Correlation parameter* $A = 0.95$, *Interval length* $N = 10$.

4.3.3 Large correlation parameter

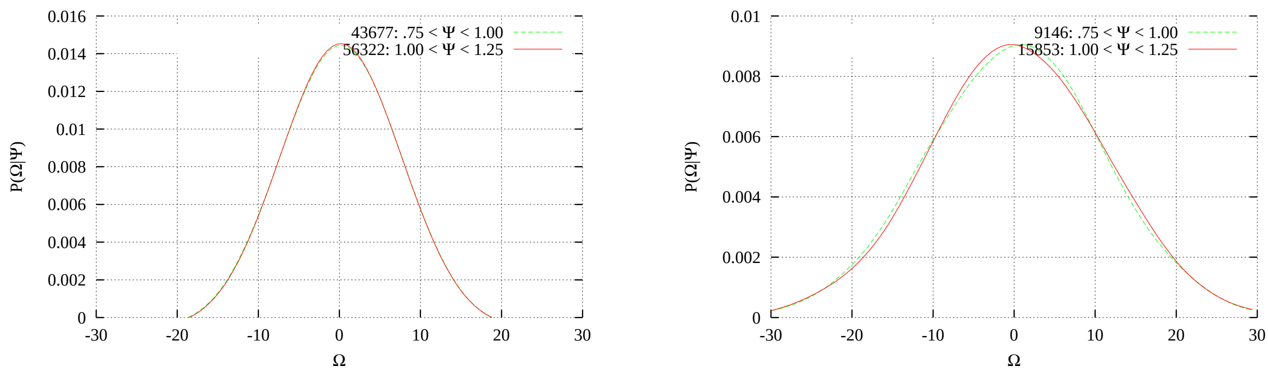
In the following experiment the correlation parameter is set to $A = 0.95$. Fig.4.6(a) and (b) show that a large correlation parameter also yields uni- and bi-modality of the probability density $P(\Omega|\Psi)$. In this experiment we sample from $N(\pm 1, 0.25)$ and $N(\pm 1, 0.5)$ respectively.

As stated in the introduction, we have reproduced the uni- and bi-modality of the probability density $P(\Omega|\Psi)$ of the demand Ω conditioned on its local noise Ψ. We can even make it appear and disappear based on the value of the correlation parameter. In the case of uncorrelated data we reproduce the uni- and bi-modality of the probability density $P(\Omega|\Psi)$ based on the value we attribute to the mean and the standard deviation of the normal distribution.

4.4 Longer intervals

In this section, we reproduce the uni- and bi-modality of the probability density $P(\Omega|\Psi)$.

The experiment consists of varying the length N of the interval. We use the same procedure as in the previous sections when performing this experiment. We consider three different values of the correlation parameter A and for each value of A we consider the case where



(a) *Interval length* $N = 100$ (b) *Interval length* $N = 400$

Figure 4.7: $P(\Omega|\Psi)$: *Correlation parameter* $A = 0.05$.

the length N of the interval is 100 and 400.

We observe that the uni- and bi-modal property of the conditional probability density $P(\Omega|\Psi)$ persists as the length N of the interval is increased.

4.4.1 Small correlation parameter

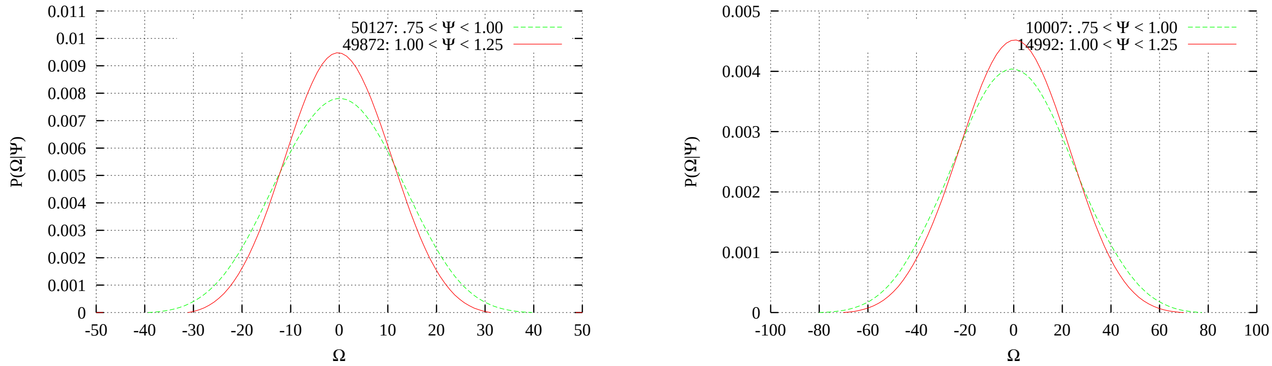
We set the correlation parameter A to 0.05. We first vary the length N of the interval to 100 and we next set the interval length N to 400 . Fig.4.7(a)and (b) shows that the probability density $P(\Omega|\Psi)$ is essentially uni-modal in the two experiments.

4.4.2 Uncorrelated data

We perform experiments with the correlation parameter $A = 0.5$ which produces uncorrelated data. The probability density $P(\Omega|\Psi)$ is uni-modal in both case of length $N = 100$ and $N = 400$ as shown in Fig.4.8(a) and (b) respectively.

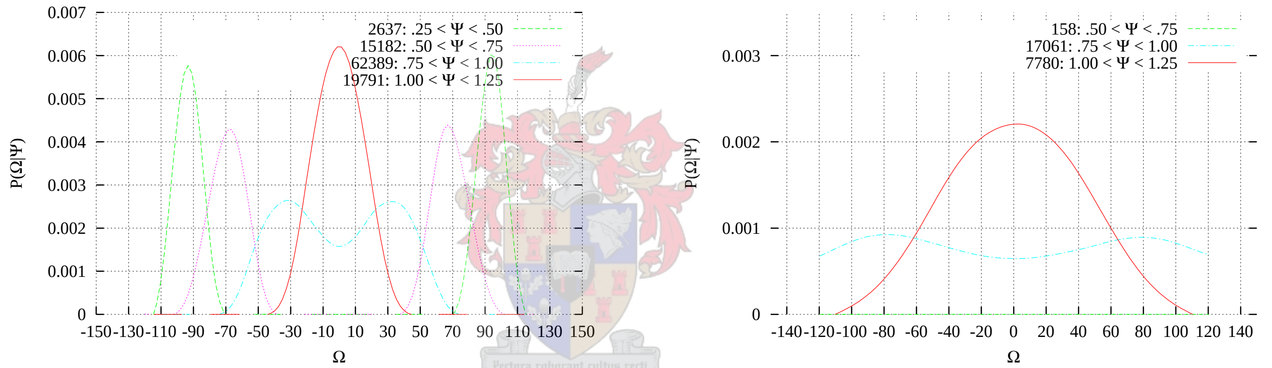
4.4.3 Large correlation parameter

We consider a large correlation parameter $A = 0.95$. The probability density $P(\Omega|\Psi)$ is uni- and bi-modal in both cases of length $N = 100$ and $N = 400$ as shown in Fig.4.9(a)



(a) *Interval length* $N = 100$ (b) *Interval length* $N = 400$

Figure 4.8: $P(\Omega|\Psi)$: *Correlation parameter* $A = 0.5$.



(a) *Interval length* $N = 100$ (b) *Interval length* $N = 400$

Figure 4.9: $P(\Omega|\Psi)$: *Correlation parameter* $A = 0.95$.

and Fig.4.9(b) respectively.

The uni- and bi-modality of the probability density $P(\Omega|\Psi)$ of Ω conditioned on Ψ persists as the number of transactions per interval is increased from 10 to 400.

4.5 Chapter summary

This chapter reproduces two-phase behaviour in a sequence of random variables. The first step we have considered is to reproduce two-phase behaviour using a sequence of random variables sampled from a Markov modulated Bernoulli process. The second step we use a variant of the Markov modulated Bernoulli process to model the behaviour of the trading shares of a single stock and reproduce two-phase behaviour. We find the existence of a critical threshold similar to that in [1]. Finally, we condider varying the interval length and reproduce the uni- and bi-modality of the conditional probability density $P(\Omega|\Psi)$.

Chapter 5

Multi modal behaviour among random variables sampled from different distributions

5.1 Introduction

In this chapter we consider the heavy tailed Weibull and Pareto distributions and reproduce multi modal behaviour among sequences sampled from these distributions.

The process of buying and selling a stock is represented by a sequence of independent and identically distributed random variables $\{X_i = a_i Y_i\}$ where $\{Y_i\}$ is sampled from a Weibull and Pareto distribution respectively and

$$a_i = a_i(Z_i) = \begin{cases} +1 & Z_i < 0.5 \\ -1 & Z_i \geq 0.5 \end{cases}$$

where Z_i is a random variable sampled from a uniform distribution $U(0, 1)$

The samples X_i are need to calculate the demand

$$\Omega = \sum_{i=1}^{N} X_i$$

where $i = 1, \ldots, N$ labels the transactions in the interval, and the local noise intensity

$$\Psi = \langle |X_i - \langle X_i \rangle| \rangle$$

where $\langle \cdots \rangle$ denotes the local expectation value computed from all transactions in the interval.

The use of market terminology is an analogy we make in order to use the concepts of demand and local noise as used in the financial market literature.

5.2 Random variables sampled from a Weibull distribution

In this section we reproduce multi-modal behaviour among sequences of random variables sampled from the heavy tailed Weibull distribution.

We generate a sequence of $M = 10^8$ random variables $\{X_i = a_i Y_i\}$ where the sequence $\{Y_i\}$ is sampled from a Weibull distribution with probability density function

$$P(x) = \gamma \alpha^{-\gamma} x^{\gamma - 1} e^{-(x/\alpha)^\gamma}$$

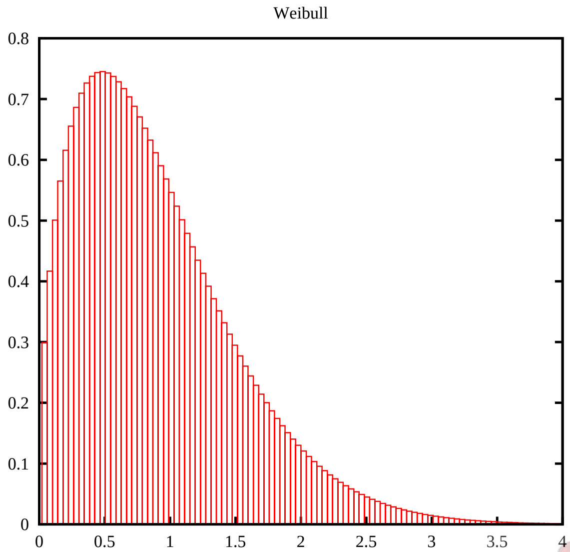
for $x \in [0, \infty)$ where α is the scale parameter and γ is the shape parameter.

We use the inversion method described above to generate a sequence $\{Y_i\}$ of random variables sampled from the Weibull distribution by transforming a continuous uniform random variable in the range 0 to 1 referred to as z with the inverse Weibull distribution function which is defined as

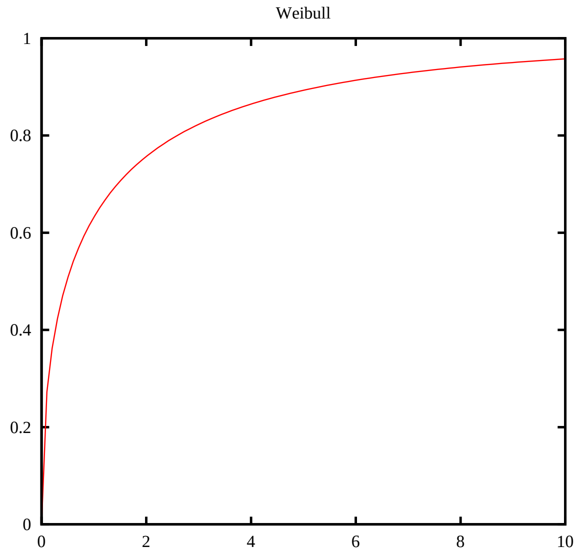
$$G(z) = \frac{\alpha}{\gamma} \log \frac{1}{1 - z}.$$

Figs.5.1 (a), (b) and (c) show respectively a graph of the Weibull probability density function, distribution function and a graph of the random variables $\{X_i = a_i Y_i\}$.

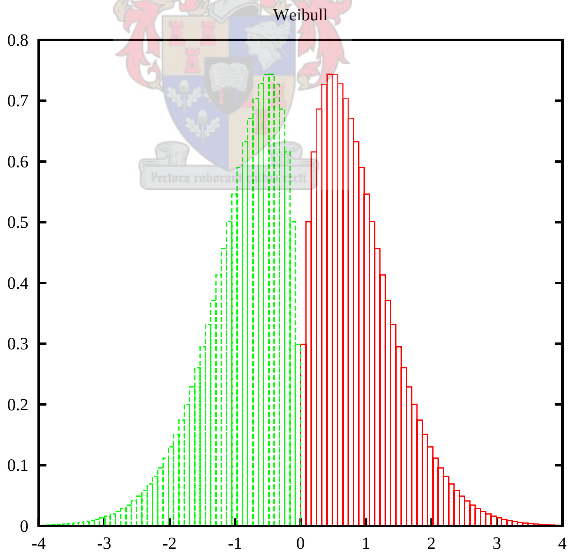
We reproduce multi modal behaviour in the conditional density $P(\Omega|\Psi)$ for different values of the shape parameter γ of the Weibull distribution. For each value of the shape parameter γ we perform experiments with different values of the interval length N.



(a) *The Weibull probability density function*



(b) *The Weibull distribution function*



(c) *The random variables* $\{X_i = a_iY_i\}$

Figure 5.1: *(a) The Weibull probability density function (b) the Weibull distribution function and (c) the random variables* $\{X_i = a_iY_i\}$ *where* $\{Y_i\}$ *is sampled from the Weibull density function.*

Algorithm 4 is used to generate a sequence $\{X_i = a_i Y_i\}$ from the Weibull distribution.

Algorithm 4

```
a = 1
initialise( A = 0.5, γ)
for i = 1 to M  do
  Z = U(0,1)          // a standard uniform RV
  if Z ≥ A then
    a = −1
  else
    a = 1
  end if
  Y = (α/γ) log 1/(1−Z)     // a Weibull RV
  X = aY
end for
```

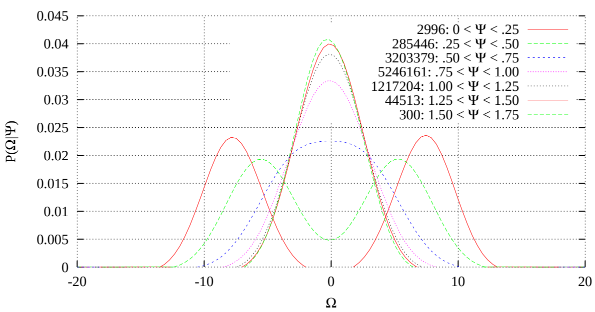
5.2.1 Shape parameter $\gamma = 2$

Figs.5.2 (a) through (e) show that for $N = 10, 20, 40, 60, 80$ the conditional probability distribution $P(\Omega|\Psi)$ is uni- and bi-modal. Fig.5.2 (f) shows that $P(\Omega|\Psi)$ is essentially uni-modal for $N = 100$.

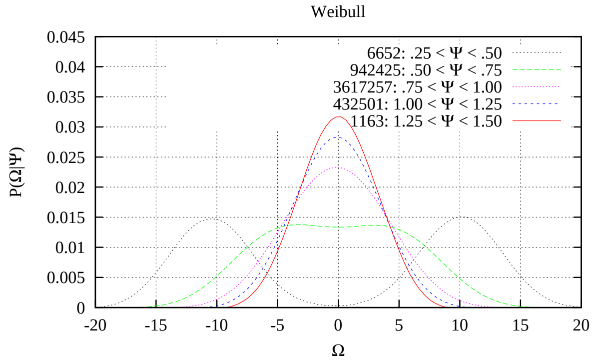
Fig.5.3 (a) through (e) show the most probable value of the demand Ω as a function of the local noise Ψ for $N = 10, 20, 40, 60, 80$. We observe the existence of a critical threshold, Ψ_c. For $\Psi < \Psi_c$, two most probable values emerge that are symmetrical around zero. For $\Psi > \Psi_c$, the most probable value of demand is approximately zero.

5.2.2 Shape parameter $\gamma = 5$

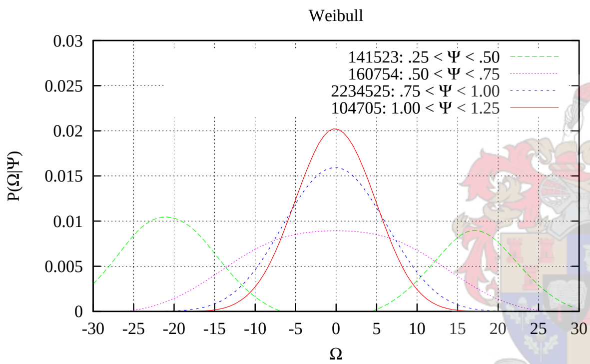
Fig.5.4 (a) through (d) show that for the interval length $N = 10, 20, 40, 60$, the conditional distribution $P(\Omega|\Psi)$ is uni- and bi-modal. Fig.5.4.(e) and (f) show that $P(\Omega|\Psi)$ is essentially uni-modal for $N = 80, 100$.



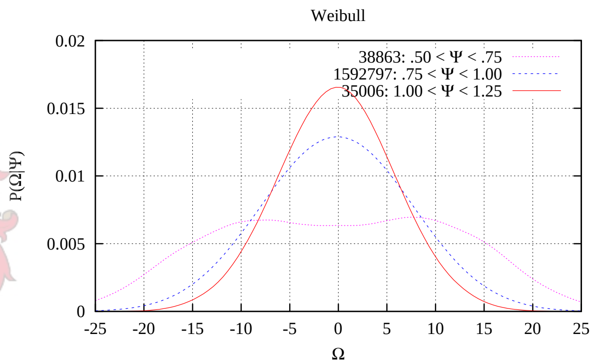
(a) *Interval length* $N = 10$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0 < \Psi \leq 0.5$ *and uni-modal for* $0.5 < \Psi \leq 1.75$



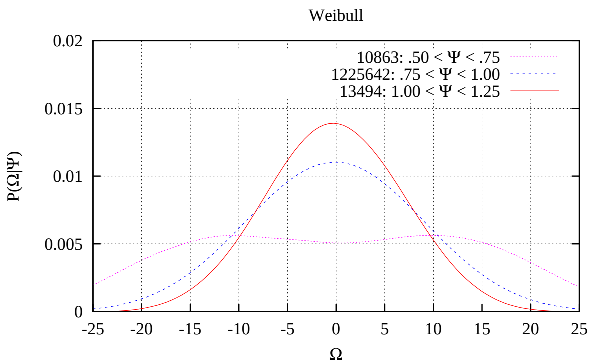
(b) *Interval length* $N = 20$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0.25 < \Psi \leq 0.75$ *and uni-modal for* $0.75 < \Psi \leq 1.25$



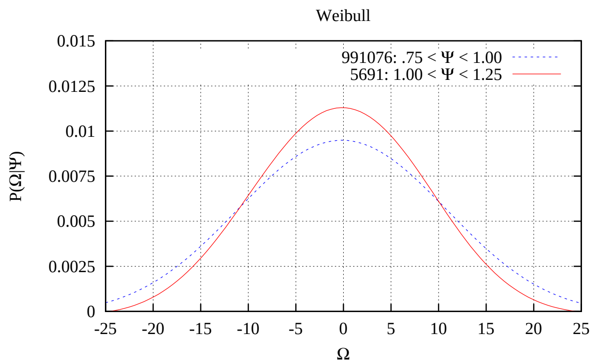
(c) *Interval length* $N = 40$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0.25 < \Psi \leq 0.5$ *and uni-modal for* $0.5 < \Psi \leq 1.25$



(d) *Interval length* $N = 60$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0.5 < \Psi \leq 0.75$ *and uni-modal for* $0.75 < \Psi \leq 1.25$

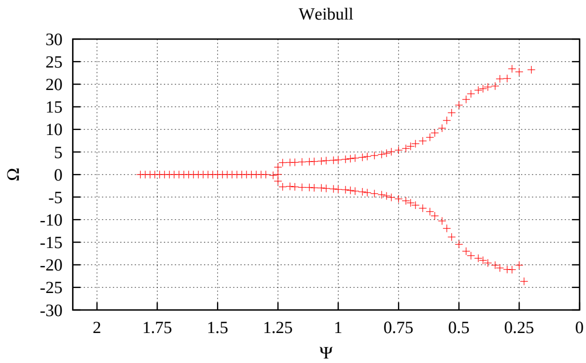


(e) *Interval length* $N = 80$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0.5 < \Psi \leq 0.75$ *and uni-modal for* $0.75 < \Psi \leq 1.25$

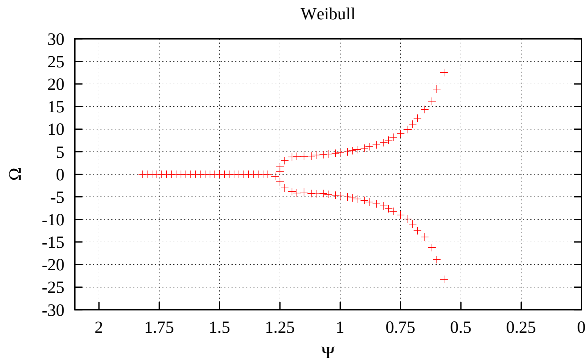


(f) *Interval length* $N = 100$ *:*$P(\Omega|\Psi)$ *is essentially uni-modal*

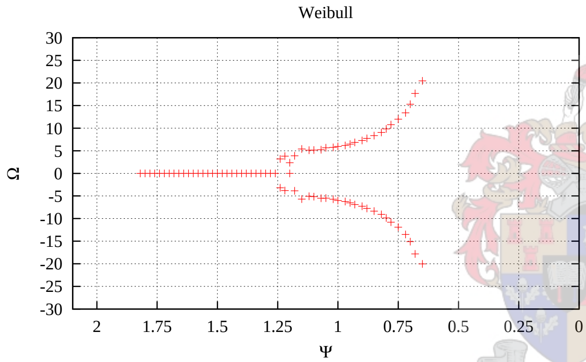
Figure 5.2: $P(\Omega|\Psi)$ *using the Weibull distribution : shape parameter* $\gamma = 2$.



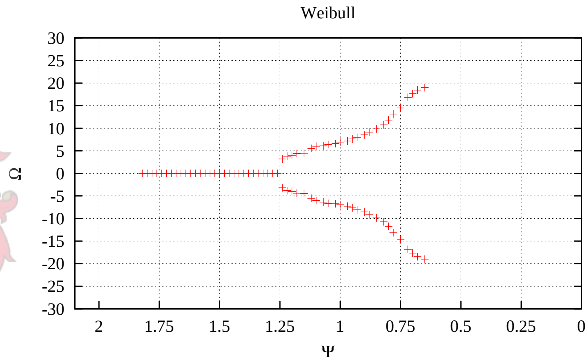
(a) *Interval length* $N = 10$



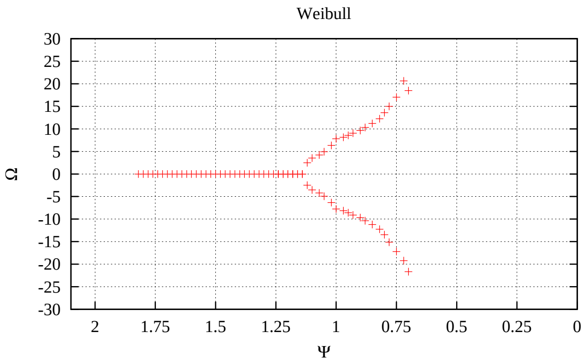
(b) *Interval length* $N = 20$



(c) *Interval length* $N = 40$

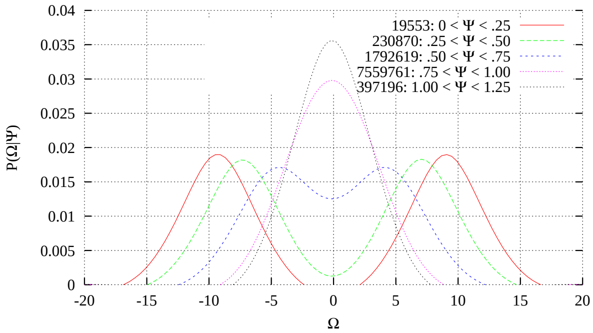


(d) *Interval length* $N = 60$

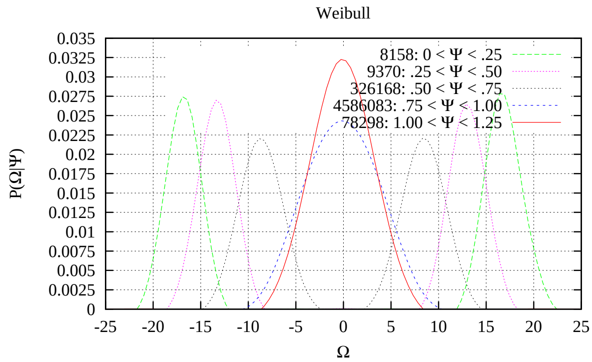


(e) *Interval length* $N = 80$

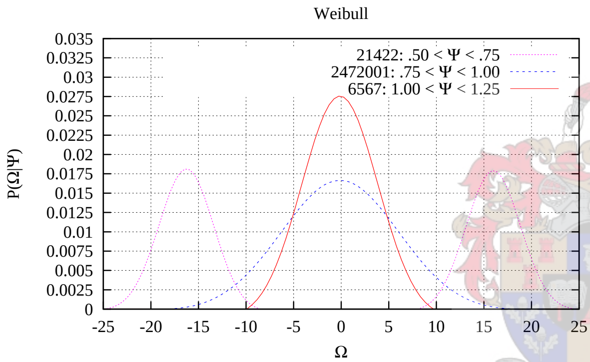
Figure 5.3: *The most probable value of the demand* Ω *as a function of the local noise* Ψ*.*



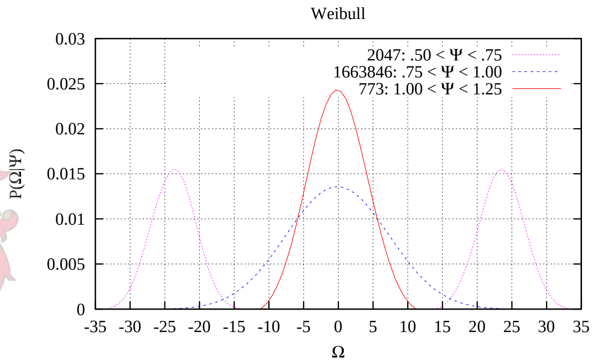
(a) *Interval length* $N = 10$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0 < \Psi \leq 0.75$ *and uni-modal for* $0.75 < \Psi \leq 1.25$



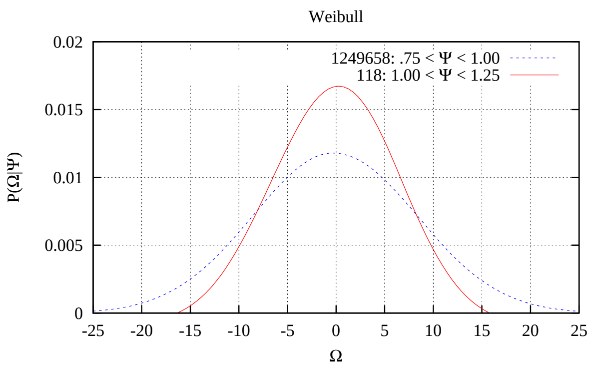
(b) *Interval length* $N = 20$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0 < \Psi \leq 0.75$ *and uni-modal for* $0.75 < \Psi \leq 1.25$



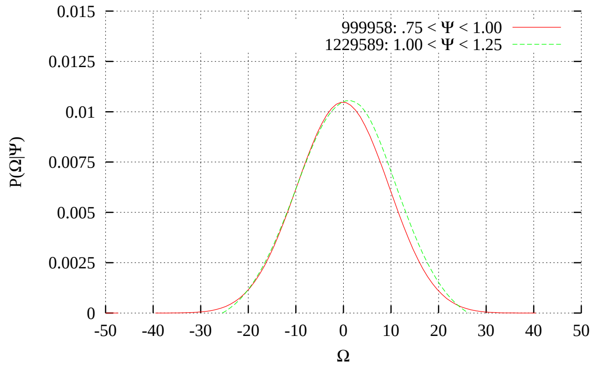
(c) *Interval length* $N = 40$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0.5 < \Psi \leq 0.75$ *and uni-modal for* $0.75 < \Psi \leq 1.25$



(d) *Interval length* $N = 60$ *:* $P(\Omega|\Psi)$ *is bi-modal for* $0.5 < \Psi \leq 0.75$ *and uni-modal for* $0.75 < \Psi \leq 1.25$



(e) *Interval length* $N = 80$ *:* $P(\Omega|\Psi)$ *is essentially uni-modal*



(f) *Interval length* $N = 100$ *:* $P(\Omega|\Psi)$ *is essentially uni-modal*

Figure 5.4: $P(\Omega|\Psi)$ *using the Weibull distribution : shape parameter* $\gamma = 5$.

5.3 Random variables sampled from a Pareto distribution

We next generate a sequence of $M = 10^8$ random variables $\{X_i = a_i Y_i\}$ where the sequence $\{Y_i\}$ is sampled from a Pareto distribution with probability density function

$$P(\gamma, x) \quad = \quad \gamma / x^{\gamma+1}$$

and

$$a_i = a_i(Z_i) = \begin{cases} +1 & Z_i < 0.5 \\ -1 & Z_i \geq 0.5 \end{cases}$$

where Z_i is a random variable sampled from a uniform distribution $U(0, 1)$.

The inverse Pareto distribution used to sample the sequence $\{Y_i\}$ from the Pareto distribution is given by

$$G(z) \quad = \quad \beta/(1 - z)^{1/\gamma}.$$

We use the inversion method described above to generate the sequence $\{Y_i\}$ from the Pareto distribution.

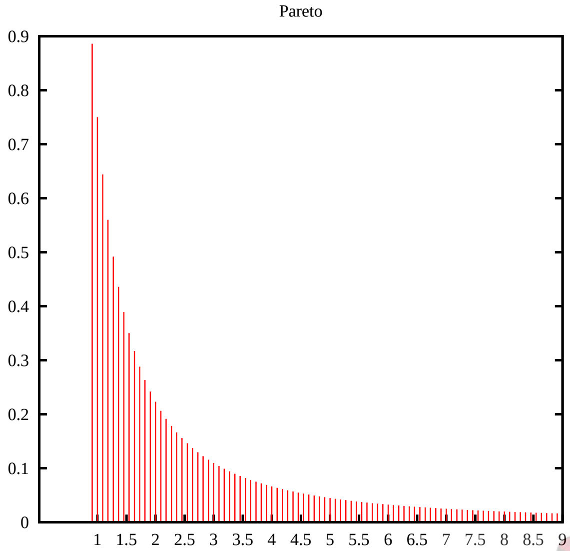
Figs.5.5 (a), (b) and (c) show the graphs of the Pareto probability density function, the Pareto distribution function and the random variables $\{X_i = a_i Y_i\}$.

Algorithm 5 is used to generate the sequence $\{X_i = a_i Y_i\}$ and we consider cases where the shape parameter $\gamma = 1.5$ and $\gamma = 15$.

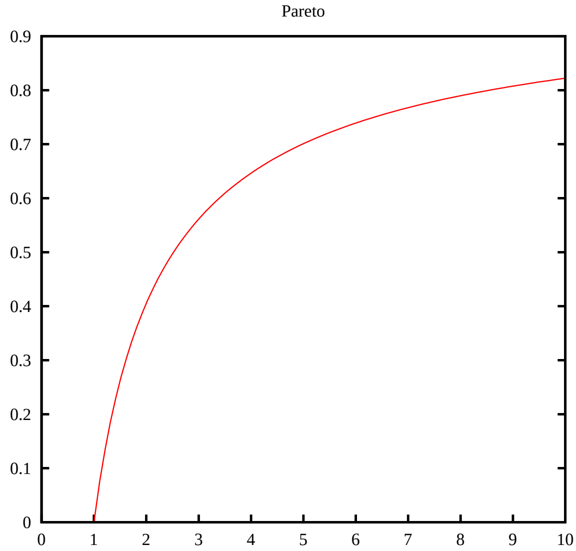
For $\gamma = 1.5$ we perform experiments with the interval length $N = 10, 20, 40, 60, 80$ and 100:

1. For $N = 10$ we observe that as the local noise intensity Ψ is increased, the probability density $P(\Omega|\Psi)$ changes from a multi-modal to a uni-modal to a bi-modal form.

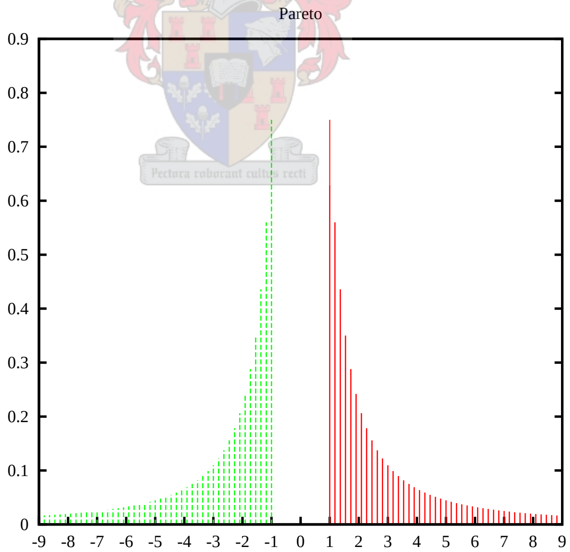
2. For $N = 20, 40, 60, 80$ we observe that as Ψ increases the probability density $P(\Omega|\Psi)$ changes from a uni-modal to a bi-modal form.



(a) *The Pareto probability density function*



(b) *The Pareto distribution function*



(c) *The random variables* $\{X_i = a_i Y_i\}$

Figure 5.5: *(a) The Pareto probability density function (b) the Pareto distribution function and (c) the random variables* $\{X_i = a_i Y_i\}$ *where* $\{Y_i\}$ *is sampled from the Pareto distribution.*

Algorithm 5

```
a = 1
initialise( A = 0.5, γ, β)
for i = 1 to M  do
  Z = U(0,1)          // a standard uniform RV
  if Z ≥ A then
    a = −1
  else
    a = 1
  end if
  Y = β/(1 − Z)^(1/γ)      // a Pareto RV
  X = aY
end for
```

3. For $N = 100$ we observe that the probability density $P(\Omega|\Psi)$ is uni-modal.

For $\gamma = 15$ we perform experiments with the interval length $N = 10$ and $N = 100$:

1. For $N = 10$ we observe that the probability density $P(\Omega|\Psi)$ is uni- and bi-modal.
2. For $N = 100$ we observe that the probability density $P(\Omega|\Psi)$ is also uni- and bi-modal.

We compute the demand Ω and the local noise intensity Ψ as stated above. The interval length N is 10. Fig.5.6 (a) shows that for $0.25 < \Psi \leq 1.5$ the probability density $P(\Omega|\Psi)$ is multi-modal.

Fig.5.6 (b) shows that $P(\Omega|\Psi)$ becomes uni-modal for $1.5 < \Psi \leq 2$. When the noise Ψ becomes larger, the probability density $P(\Omega|\Psi)$ becomes bi-modal as shown in Fig.3.6 (c).

5.3.1 Longer intervals

We perform experiments with different values of the interval length N. The shape parameter γ is 1.5.

We first set the interval length $N = 20$. Fig.5.7 (a) shows that $P(\Omega|\Psi)$ is bi-modal for $0.75 < \Psi \leq 1.5$. Fig.5.7 (b) shows that $P(\Omega|\Psi)$ is uni-modal for $1.5 < \Psi \leq 3$ and Fig.5.7 (c) shows that $P(\Omega|\Psi)$ is uni- and bi-modal for $3 < \Psi \leq 6.25$.

We next set the interval length N to 40. Fig.5.8 (a) shows that $P(\Omega|\Psi)$ is uni-modal for $1.5 < \Psi \leq 3.5$ and Fig.5.8 (b) shows that $P(\Omega|\Psi)$ is bi-modal for $3.75 < \Psi \leq 6.25$.

We then set the interval length N to 60. Fig.5.9 (a) shows that $P(\Omega|\Psi)$ is uni-modal for $1.75 < \Psi \leq 3.25$ and Fig.5.9 (b) shows that $P(\Omega|\Psi)$ is bi-modal for $3.5 < \Psi \leq 4.25$.

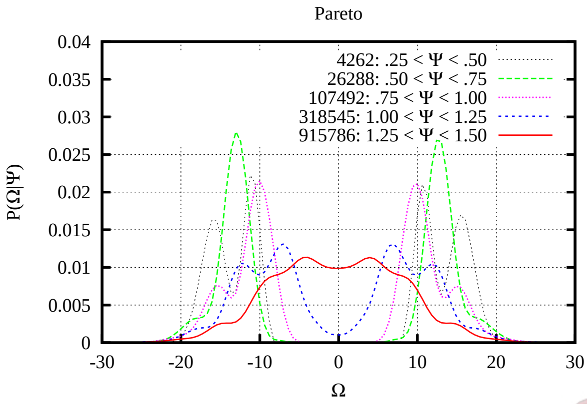
We next set the interval length N to 80. Fig.5.10 (a) shows that $P(\Omega|\Psi)$ is uni-modal for $1.75 < \Psi \leq 3.5$ and Fig.5.10 (b) shows that $P(\Omega|\Psi)$ is bi-modal for $3.5 < \Psi \leq 3.75$.

We finally set the interval length N to 100. Fig.5.11 shows that the probability density is essentially uni-modal for $1.75 < \Psi \leq 3.75$.

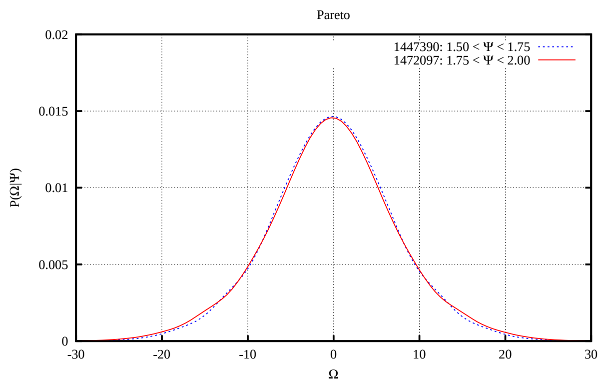
Fig.5.12 (a), (b), and (c) shows respectively the graph of the most probable value of the demand Ω as a function of the local noise Ψ for the interval length $N = 10, 20$ and $N = 40$. The shape parameter γ is 1.5.

5.3.2 Increasing the value of the shape parameter γ

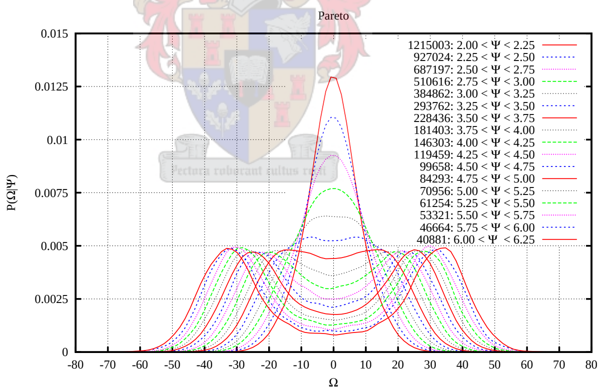
We next increases the shape parameter γ to 15 and we reproduce bimodal behaviour when the interval length $N = 10$ and $N = 100$. Fig.5.13(a) shows the case of interval length $N = 10$ and Fig.5.13(b) shows the case of interval length $N = 100$.



(a) $0.25 < \Psi \leq 1.5$: *the probability density* $P(\Omega|\Psi)$ *is multi-modal*

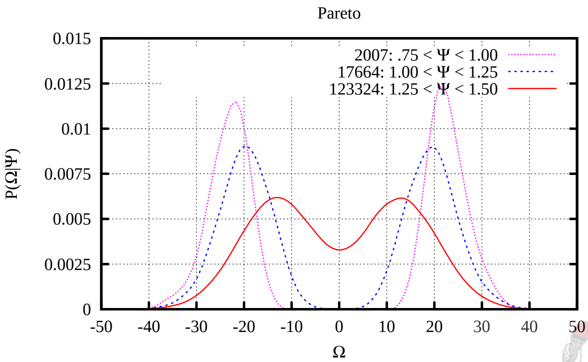


(b) $1.5 < \Psi \leq 2$: *the probability density* $P(\Omega|\Psi)$ *is uni-modal*

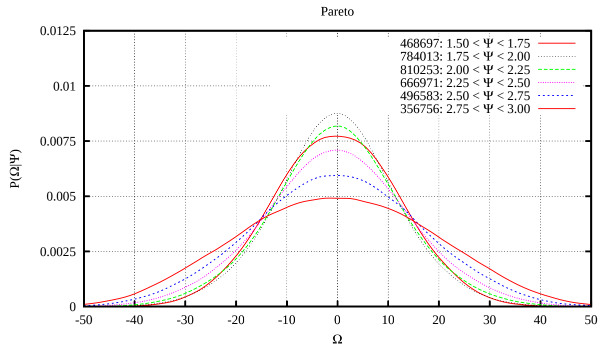


(c) $2 < \Psi \leq 3.25$: *the probability density* $P(\Omega|\Psi)$ *is uni-modal.* $3.25 < \Psi \leq 6.25$: *the probability density* $P(\Omega|\Psi)$ *is bi-modal*

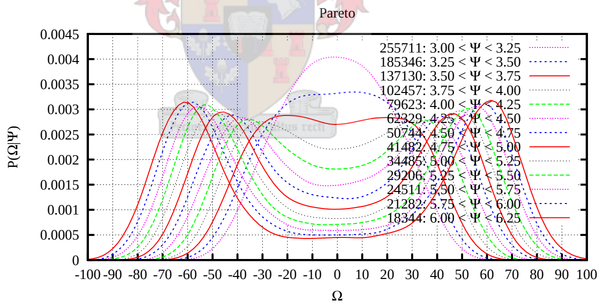
Figure 5.6: $P(\Omega|\Psi)$ *using the Pareto distribution : shape parameter* $\gamma = 1.5$, *interval length* $N = 10$.



(a) $0.75 < \Psi \leq 1.5$: *the probability density* $P(\Omega|\Psi)$ *is bi-modal*

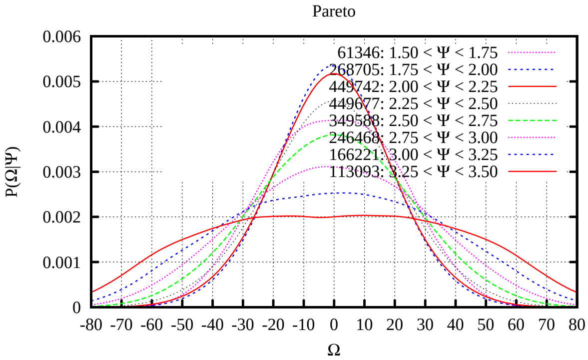


(b) $1.5 < \Psi \leq 3$: *the probability density* $P(\Omega|\Psi)$ *is uni-modal*

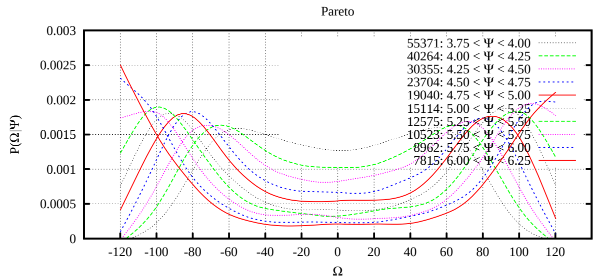


(c) $3 < \Psi \leq 6.25$: *the probability density* $P(\Omega|\Psi)$ *is uni- and bi-modal*

Figure 5.7: *$P(\Omega|\Psi)$ using the Pareto distribution : shape parameter $\gamma = 1.5$, interval length $N = 20$.*

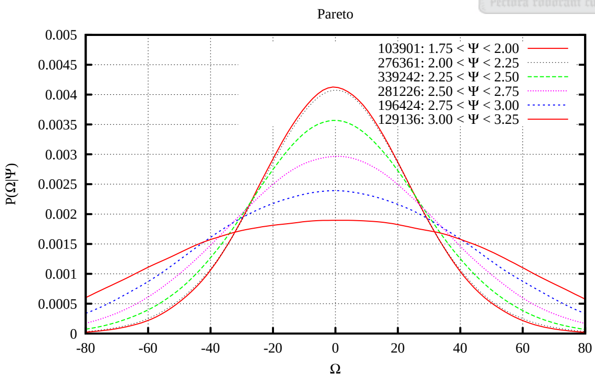


(a) $1.5 < \Psi \leq 3.5$: *the probability density* $P(\Omega|\Psi)$ *is uni-modal*

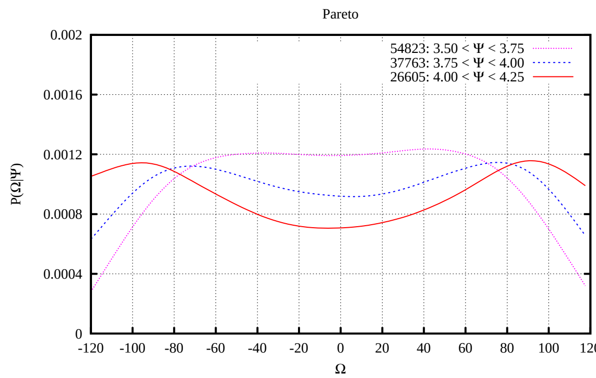


(b) $3.75 < \Psi \leq 6.5$: *the probability density* $P(\Omega|\Psi)$ *is bi-modal*

Figure 5.8: $P(\Omega|\Psi)$ *using the Pareto distribution : shape parameter* $\gamma = 1.5$ *, interval length* $N = 40$.

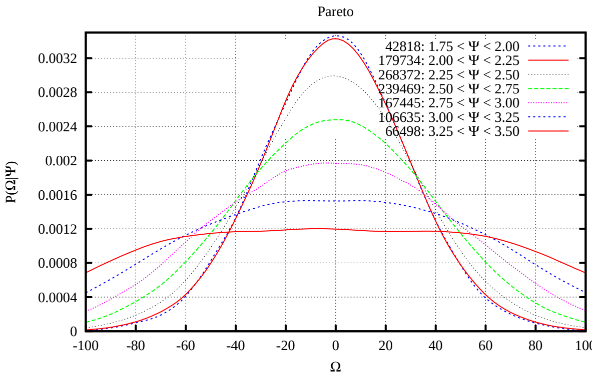


(a) $1.75 < \Psi \leq 3.25$: *the probability density* $P(\Omega|\Psi)$ *is uni-modal*

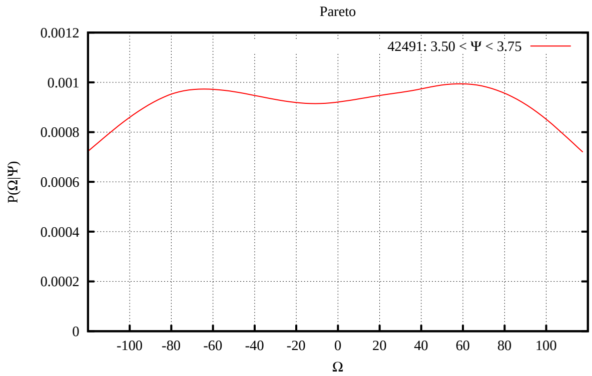


(b) $3.5 < \Psi \leq 4.25$: *the probability density* $P(\Omega|\Psi)$ *is bi-modal*

Figure 5.9: $P(\Omega|\Psi)$ *using the Pareto distribution: shape parameter* $\gamma = 1.5$ *, interval length* $N = 60$.



(a) $1.75 < \Psi \leq 3.5$: *the probability density* $P(\Omega|\Psi)$ *is uni-modal*



(b) $3.5 < \Psi \leq 3.75$: *the probability density* $P(\Omega|\Psi)$ *is bi-modal*

Figure 5.10: $P(\Omega|\Psi)$ *using the Pareto distribution: shape parameter* $\gamma = 1.5$, *interval length* $N = 80$.

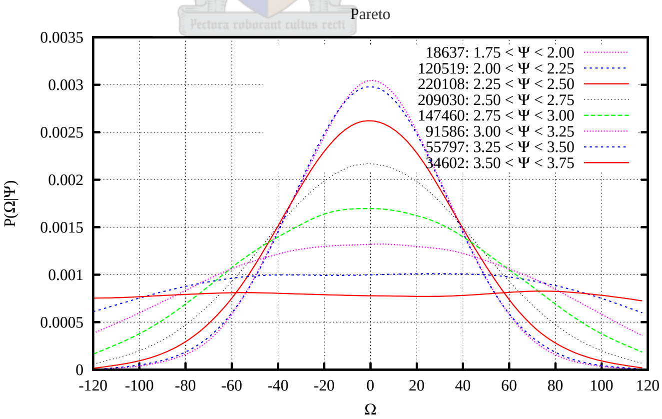
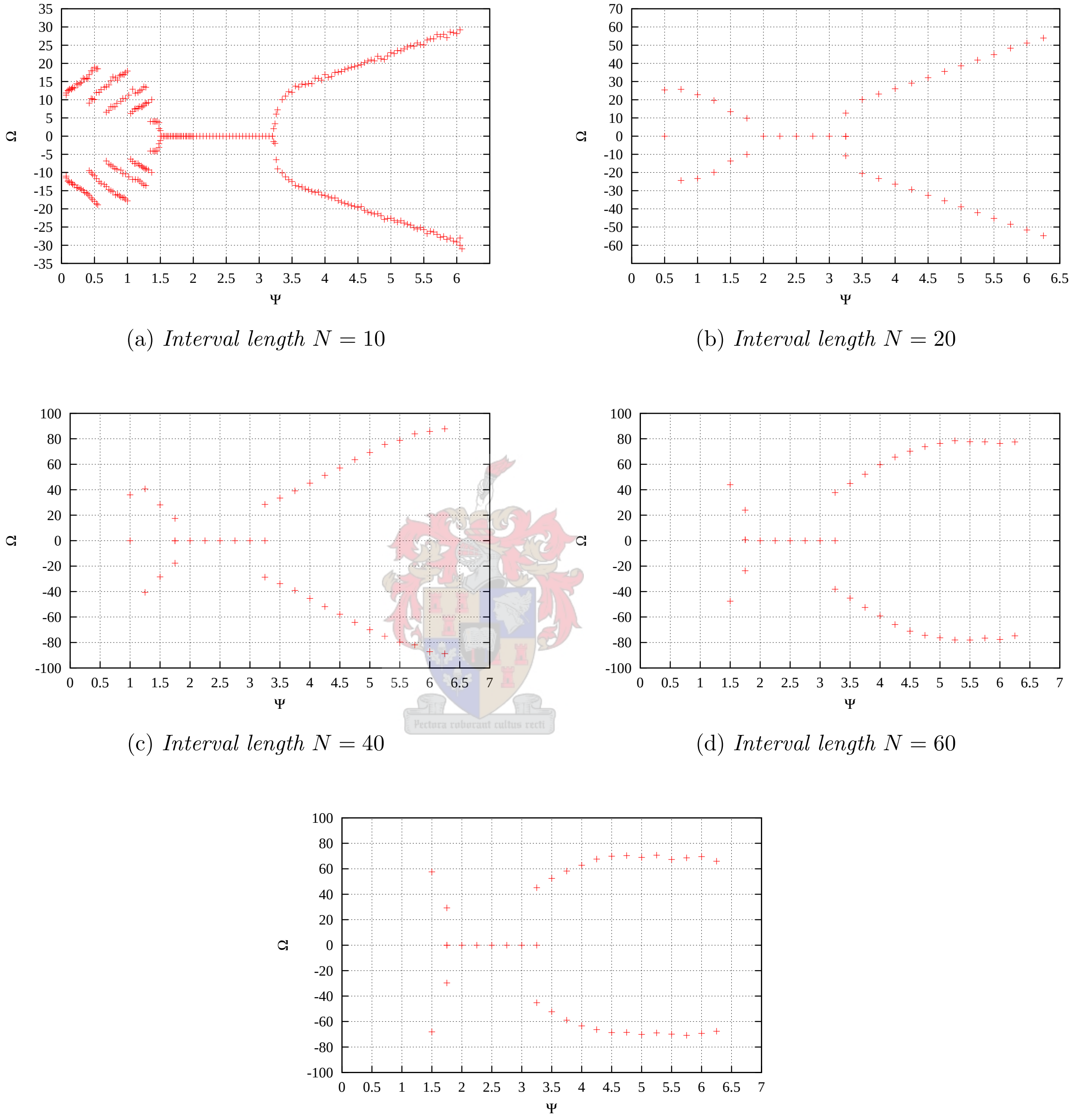


Figure 5.11: $P(\Omega|\Psi)$ *using the Pareto distribution : shape parameter* $\gamma = 1.5$, *interval* $N = 100$. *The probability density* $P(\Omega|\Psi)$ *is essentially uni-modal.*



(a) *Interval length* $N = 10$

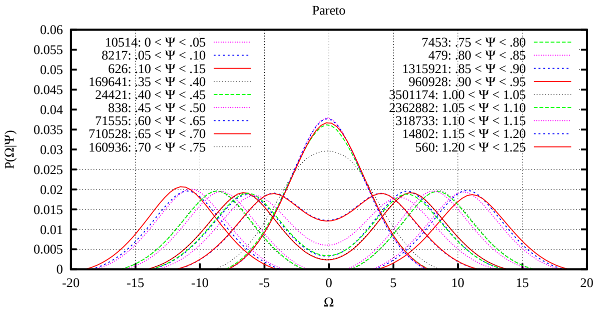
(b) *Interval length* $N = 20$

(c) *Interval length* $N = 40$

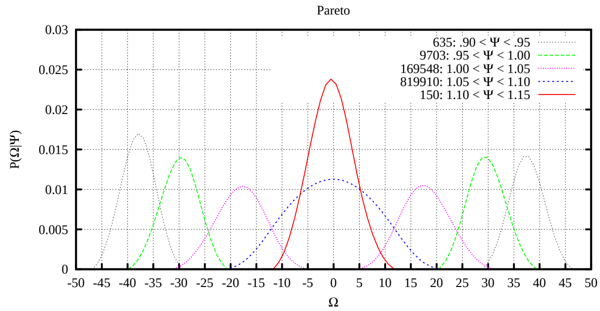
(d) *Interval length* $N = 60$

(e) *Interval length* $N = 80$

Figure 5.12: *The most probable value of the demand* Ω *as a function of the local noise* Ψ.



(a) *Interval length $N = 10$: the probability density $P(\Omega|\Psi)$ is uni- and bi-modal*



(b) *Interval length $N = 100$: the probability density $P(\Omega|\Psi)$ is uni- and bi-modal*

Figure 5.13: *$P(\Omega|\Psi)$ using the Pareto distribution : shape parameter $\gamma = 15$.*

5.4 Processing the data

We generate a sequence of $M = 10^8$ random variables $\{X_i = a_i Y_i\}$ where the sequence $\{Y_i\}$ is sampled either from a Weibull or a Pareto distribution as mentioned in the previous sections. The sequence $\{X_i = a_i Y_i\}$ is saved in a trace file. We set the interval length N and use an awk script to process the trace file to compute the demand Ω and the local noise Ψ in each interval. The sequence of Ω and Ψ values is written to a summary file.

We set the width of the demand Ω and local noise Ψ bins and use several awk scripts to compute an approximation of the conditional probability density $P(\Omega|\Psi)$. A bash script calls a gawk script several times, each time computing the probability density $P(\Omega|\Psi)$ for a new range of Ψ values: $\Psi_- \leq \Psi < \Psi_+$. This procedure produces three output files

1. A file which records $P(\Omega|\Psi)$ for each range $\Psi_- \leq \Psi < \Psi_+$.
2. A file which contains a list of gnuplot commands to plot graphs of the conditional probability density $P(\Omega|\Psi)$ for several values of the local noise Ψ.
3. A file which contains data to produce a plot of the most probable value of the demand Ω as a function of the local noise Ψ.

The Savitzky-Golay [25] smoothing filter is used to process the data in the third file mentioned above. A awk script is used to extract a column of data produced as described above. The column is processed with the Savitzky-Golay filter and another awk script is used to replace the column of raw data in the file with the smoothed data.

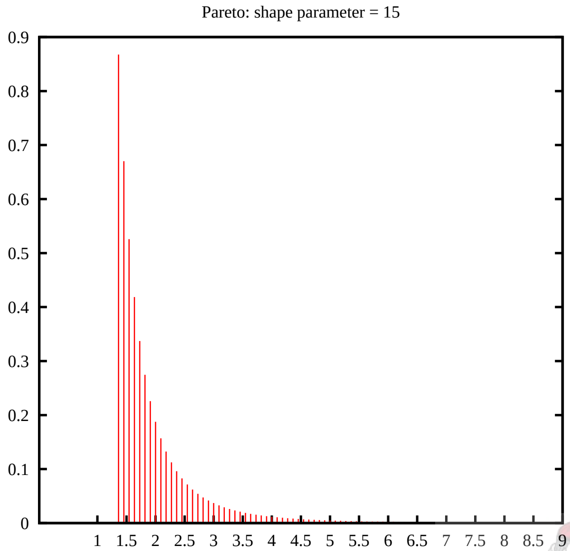
5.5 The origin of the multi-modality

The multi-modality that occurs when processing data from the Pareto distribution is the effect of the distribution tail, i.e., the rare events, which even with small probability of occurrence can significantly affect the values of the moment based descriptors calculated from a sample of random numbers.

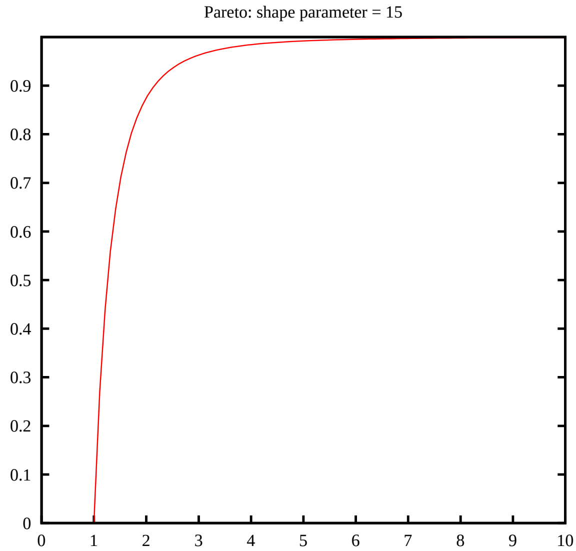
A random variable sampled from a Pareto distribution can have extreme values. The presence of the extreme values in the process causes the multi-modality that we observe in Fig.5.12 (a). We observe that the probability density $P(\Omega|\Psi)$ is uni- and bi-modal when using random variables sampled from the Pareto distribution with shape parameter $\gamma = 15$ as shown in Fig.5.13. In case of shape parameter $\gamma = 15$ Figs.5.14 (a), (b) and (c) show the graphs of the Pareto probability density function, the Pareto distribution function and the random variables $\{X_i = a_i Y_i\}$. We can observe a difference of tails between the Pareto probability density function with shape parameter $\gamma = 1.5$ as shown in Fig.5.5(a) and the Pareto probability density function with shape parameter $\gamma = 15$ as shown in Fig.5.14(a).

$P(\Omega|\Psi)$ is multi-modal in experiment performed with random variables sampled from the Pareto distribution with shape parameter $\gamma = 1.5$, which is the long tail Pareto distribution.

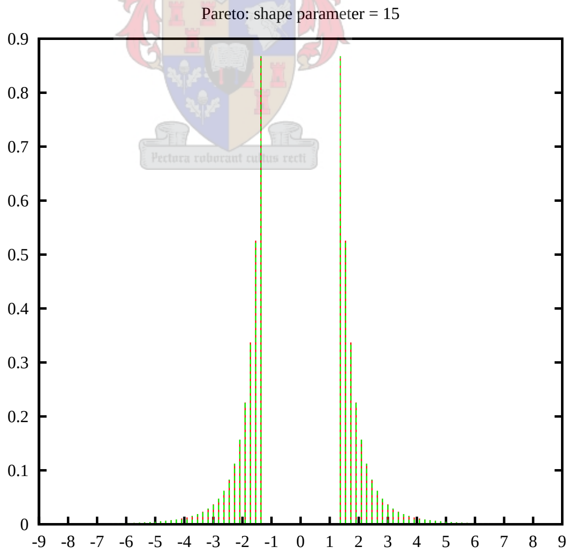
Simulation of heavy-tailed distributions for estimation of measures is time consuming, as the simulation must run exceptionally long in order to capture the effect of the distribution tail.



(a) *The Pareto probability density function*



(b) *The Pareto distribution function*



(c) *The random variables* $\{X_i = a_iY_i\}$

Figure 5.14: *The Pareto probability density function (a), the Pareto distribution function(b) and the random variables* $\{X_i = a_iY_i\}$ *where* $\{Y_i\}$ *is sampled from the Pareto distribution.*

5.6 Chapter summary

The goal of this chapter was to reproduce two-phase behaviour using random variables sampled form the Weibull and Pareto distributions. The Weibull distribution exhibits bi-modal behaviour. Fig.5.3 illustrates the existence of a critical threshold, Ψ_c. For $\Psi < \Psi_c$, two most probable values emerge that are symmetrical around zero. For $\Psi > \Psi_c$, the most probable value of demand is roughly zero. We observe that for the Weibull distribution the bifurcation occurs for $\Psi < \Psi_c$, in contrast with what Plerou [1] has found in the NYSE data, which has the bifurcation occurring for $\Psi > \Psi_c$.

Next we reproduce the multi-modal behaviour among sequences of random variables sampled from the Pareto distribution. Considering the Pareto distribution in the high noise region $2 < \Psi < 6$, the bifurcation shown in Fig.5.12 corresponds to what Plerou found in the NYSE data. Section four describes how we process the data. Finally, we explain the origin of the multi-modality.

Our procedure to reproduce two phase behaviour suggests that the trade volumes observed in the NYSE data has a heavy tailed distribution.

Chapter 6

Thesis summary

The thesis begins by presenting the problem to be addressed and introduces and motivates the issues to be explored in this thesis.

Second, we describe the phenomena of phase transitions in physics. We present the phase transition in physical systems since it has been (incorrectly) stated that phase transitions are associated with financial markets and it is a key to understanding chapter 4 and 5 of our work. We present a literature survey of two-phase behaviour of financial markets encountered in [1]. The literature survey indicates how other authors have reproduced the two-phase behaviour.

Third, we present the mathematical concepts that we use in our experiments. We begin by presenting a survey of random variables. We then introduce the Markov modulated Bernoulli process, the Weibull and the Pareto distributions. Finally, we discuss the inverse transform technique that we use to generate random variables from a probability distribution.

Fourth, we perform experiments and reproduce two-phase behaviour using sequences of random variables sampled from a Bernoulli process, a Markov modulated Bernoulli process and from a mixture of two normal distributions. The process consists of performing experiments with correlated and uncorrelated sequence of random variables. We reproduce the two-phase behaviour with large and small values of the correlation parameter. Different values of the mean and standard deviation of a normal distribution are used in our experiments. We perform experiments varying the length of the interval while keeping the other parameters fixed, and find that the uni- and bi- modality of the conditional probability density $P(\Omega|\Psi)$ occur for various values of the interval length.

Plerou [1] suggested that the two phase behaviour indicates a link between the dynamics of a financial market with many interacting participants and the phenomenon of phase transitions that occurs in physical systems with many interacting units. We show that what Plerou [1] described as phase transition in a real market can also be obtained using uncorrelated random variables sampled from probability distributions.

Fifth, we perform experiments and reproduce two phase behaviour using sequences of random variables sampled from the Weibull distribution. Two different values of the shape parameter of the Weibull distribution are used in our experiments to reproduce two- phase behaviour. $P(\Omega|\Psi)$ exhibits uni- and bi-modal behaviour. We also find the existence of a critical threshold, Ψ_c. For $\Psi < \Psi_c$, two most probable values emerge that are symmetrical around zero. For $\Psi > \Psi_c$, the most probable value of demand is roughly zero. We observe that for the Weibull distribution the bifurcation occurs for $\Psi < \Psi_c$, in contrast with what Plerou [1] found in the NYSE data, which has the bifurcation occurring for $\Psi > \Psi_c$.

Finally, we perform experiments and reproduce two-phase behaviour using sequences of random variables sampled from the Pareto distribution. Different values of the interval length are considered in our experiments. Different values of the shape parameter are used to reproduce two-phase behaviour. $P(\Omega|\Psi)$ is multi-modal in the low noise region and bi-modal in the high noise region. The origin of the multi-modality is the effect of extreme values of the Pareto distribution. The bifurcation that occurs in the high noise region corresponds to what Plerou found in the NYSE data. This would suggested that the trade volumes observed in the NYSE data has a heavy tailed distribution.

Many authors have commented on the two-phase behaviour that Plerou [1] reported. Those authors have used various methods to reproduce the two-phase behaviour. Different explanations from different authors where given about what causes the two-phase behaviour of financial markets.

Our work shows that the data sampling procedure is the cause of the two-phase behaviour.

Bibliography

[1] V. Plerou, P. Gopikrishnan and H.E. Stanley. Two-phase behaviour of financial markets. Nature Vol. 421 (9 Jan 2003) p 130.

[2] F.E. Banks. Global Finance and Financial Markets: A Modern Introduction. World Scientific Publishing Company, 2001

[3] N. Goldenfeld. Lectures on Phase Transitions and the Renormalization Group. Perseus Publishing, 1992.

[4] D.J.C. Mackay. Information Theory , Inference, and learning Algorithms. Cambridge University Press, 2004.

[5] J. Cardy. Scaling and Renormalization in Statistical Physics. Cambridge University Press, 1996.

[6] Phase transition. Wikipedia, http://en.wikipedia.org/wiki/Phase_transition.

[7] H.D. Henderson. Supply and Demand, Kessinger Publishing, 2004.

[8] K. Matia and K. Yamasaki. Statistical Properties of Demand Fluctuations in the Financial Market. Quant. Finance, 2005, 5, pp. 513 - 517.

[9] M. Potters and J.–P. Bouchard. Comment on "Two-phase behaviour of financial markets". eprint: cond-mat/0304514, 2003.

[10] S. Sinha and S. Raghavendra. Emergence of two-phase behaviour in markets through interaction and learning in agents with bounded rationalty. Proc.3rd Nikkei Econphysics symposium, Tokyo, Japan, (Nov 2004), Springer.

[11] B. Zheng, T. Qiu and F. Ren. Two-phase phenomena, minority games, and herding models. Physical Review E, Vol. 69, Issue 4, id. 046115 04/2004

[12] M. Wyart and J–P. Bouchaud. Self-referential behaviour, overreaction and conventions in financial markets. e-print cond-mat/0303584

[13] F.F. Gong, F.X. Gong and F.Y. Gong. Open dynamic behaviour of financial markets. The European Physical Journal B, Vol. 49, Issue 3, February I 2006, pp.267-268.

[14] S.J. Grossman and J.E. Stiglitz. On the Impossibility of Informationally Efficient Markets. The American Economic Review, Vol. 70, No. 3 (Jun., 1980), pp. 393-408.

[15] M. Forster and B. Halfpap. Econophysics: A simple explanation of two-phase behaviour. http://philosophy.wics.edu/forster/papers/Econopysics6.pdf

[16] A. Costa, A.E. Krzesinski, M. Ramakrishnan and P.G. Taylor. A comment on two-phase behaviour of financial markets. To appear in the Journal of Theoretical and Applied Finance, February 2007.

[17] J. Banks, J.S. Carson, B.L. Nelson and D.M. Nicol. Discrete-Event System Simulation. Prentice Hall International Series in Industrial and Systems Engineering, Prentice Hall, 2001.

[18] C.W. Helstrom. Probability and Stochastic Processes for Engineers. MacMillan Publishing Company, 1984.

[19] M. Capinski and P.E. Kopp. Measure, Integral and Probability, Second Edition. Springer-Verlag, 2003.

[20] E. Weisstein. Measurable Space. MathWorld, http://mathworld.wolfram.com/MeasurableSpace.html.

[21] Bernoulli sequence. Wikipedia, http://en.wikipedia.org/wiki/Bernoulli_process.

[22] S.Őzekici. Business and Economics and Mathematics and Statistics. Mathematical Methods of Operations Research Vol. 45, Number 3 October, 1997.

[23] R. Nelson. Probability, Stochastic Processes and Queueing Theory. Springer-Verlag, 1995.

[24] K. Sigman. A primer on heavy-tailed distributions. Queueing Systems, 33:261-275, 1999.

[25] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery. Numerical Recipes in C++, The Art of Scientific Computing, Second Edition. Cambridge University Press, 2002.